Dynameomics: Protein Mechanics, Folding and Unfolding through Large Scale All-Atom Molecular Dynamics Simulations

INCITE 6

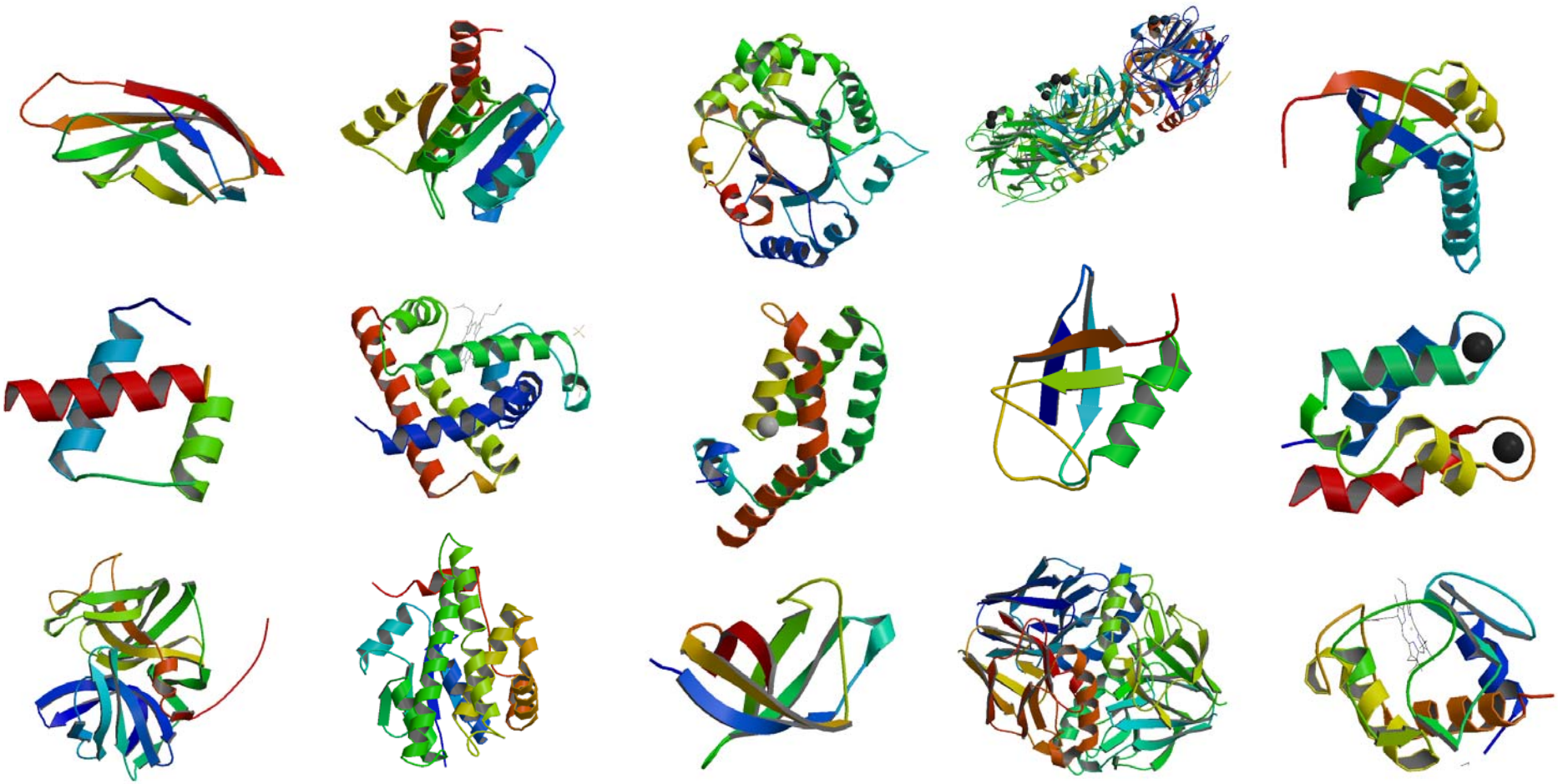
David A. C. Beck
Valerie Daggett Research Group

Department of Medicinal Chemistry
University of Washington, Seattle

November 15th, 2005

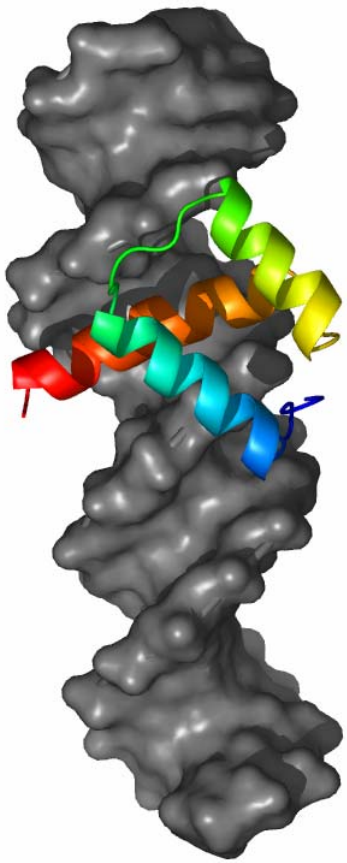
Proteins

- **Proteins are life's machines, tools and structures**
  - **Many jobs, many shapes, many sizes**

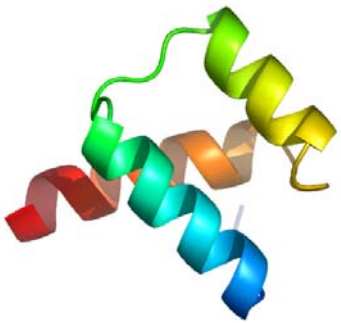


Proteins

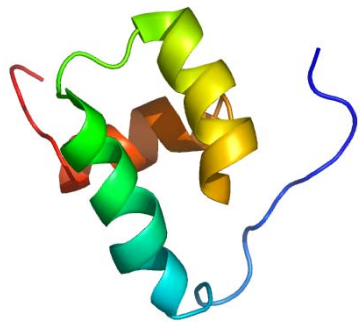
- **Proteins are life's machines, tools and structures**
  - **Nature reuses designs for similar jobs**



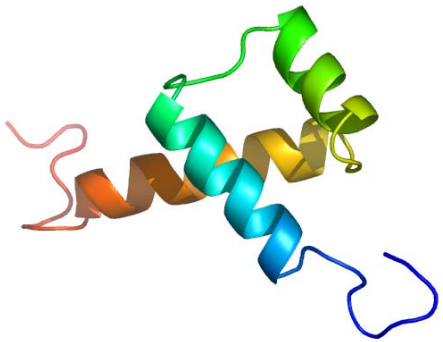
1hdd



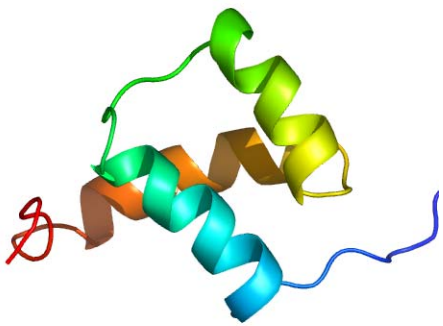
1enh



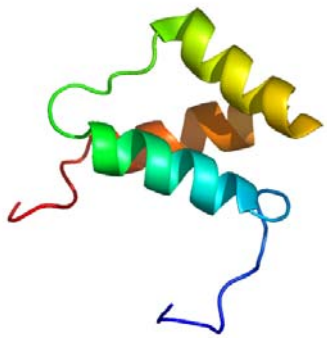
1f43



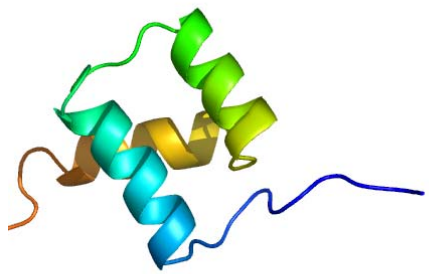
1ftt



1bw5



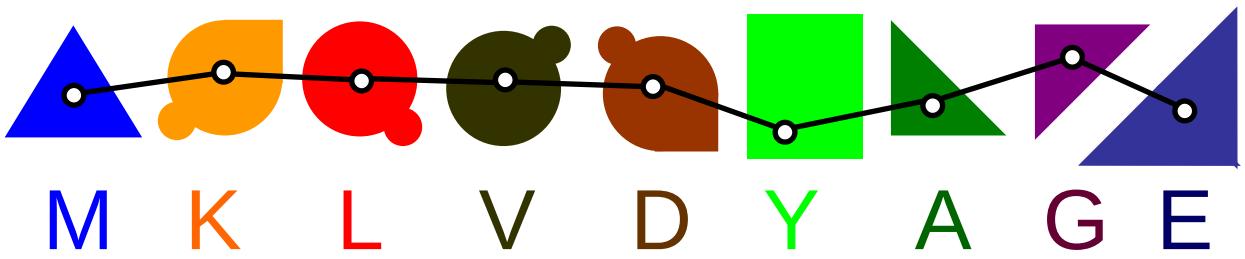
1du6



1cqt

Proteins

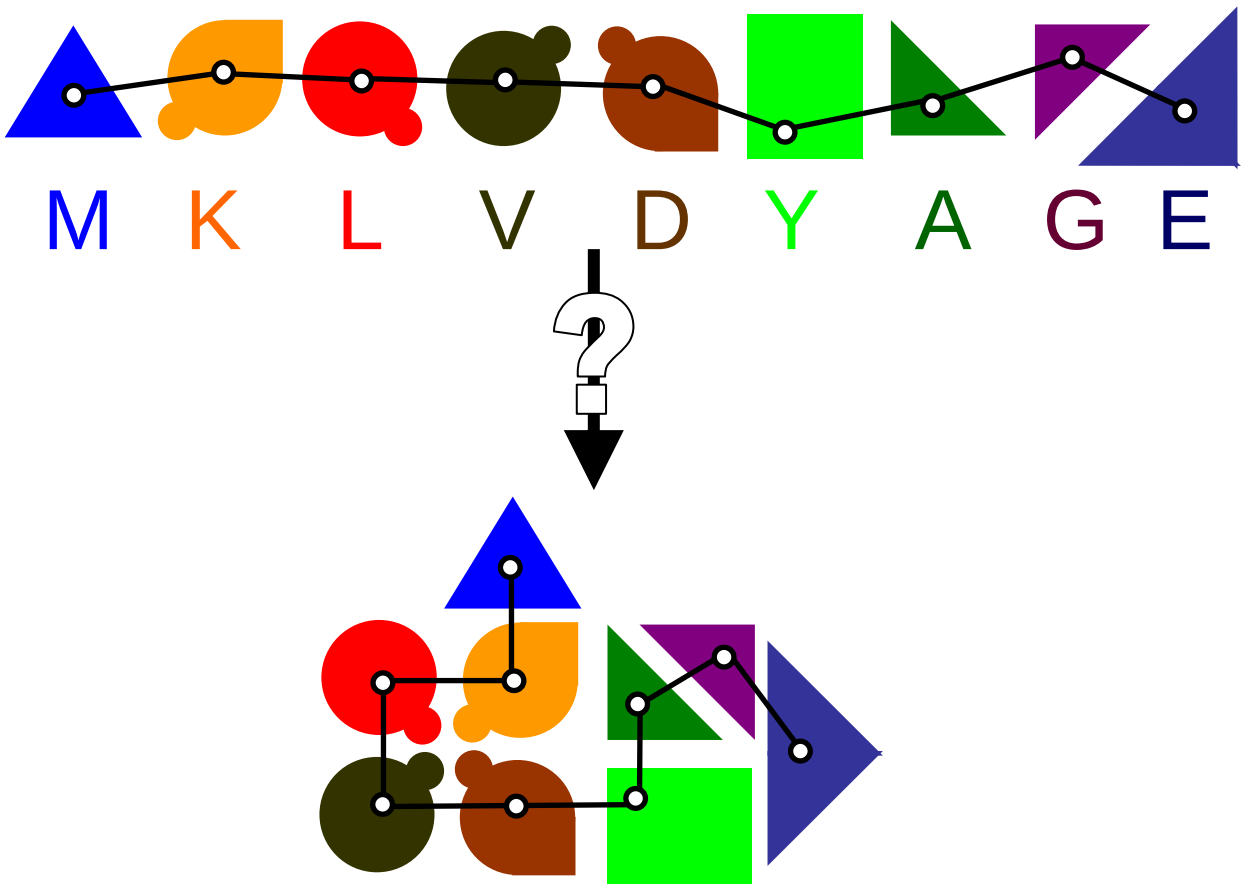
- **Proteins are hetero-polymers of specific sequence**



  - **There are 20 common polymeric units (amino acids)**
    - **Composed of a variety of basic chemical moieties**
  - **Chain lengths range from 40 amino acids on up**

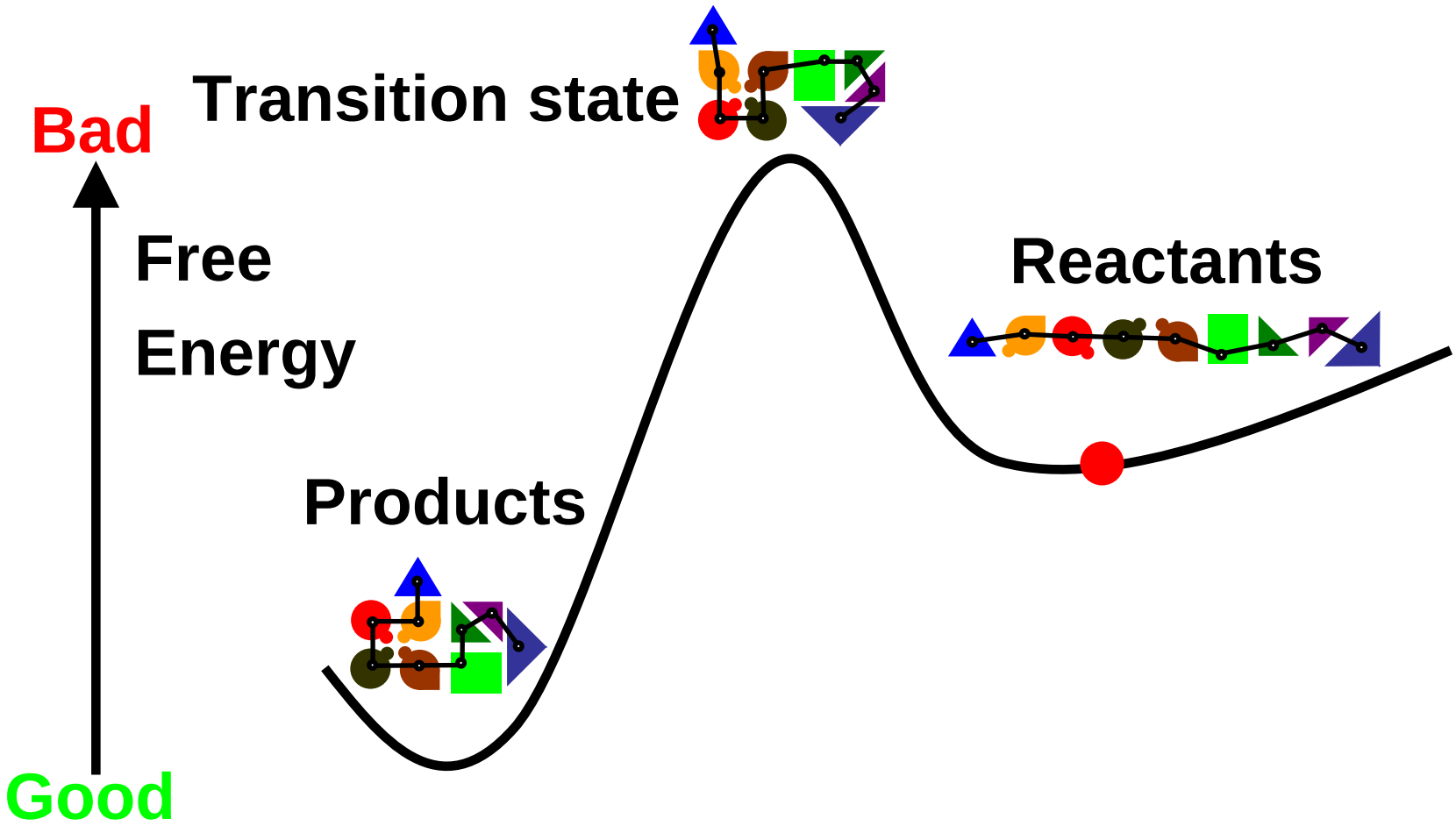
Proteins

- **Proteins are hetero-polymers that adopt a unique fold**



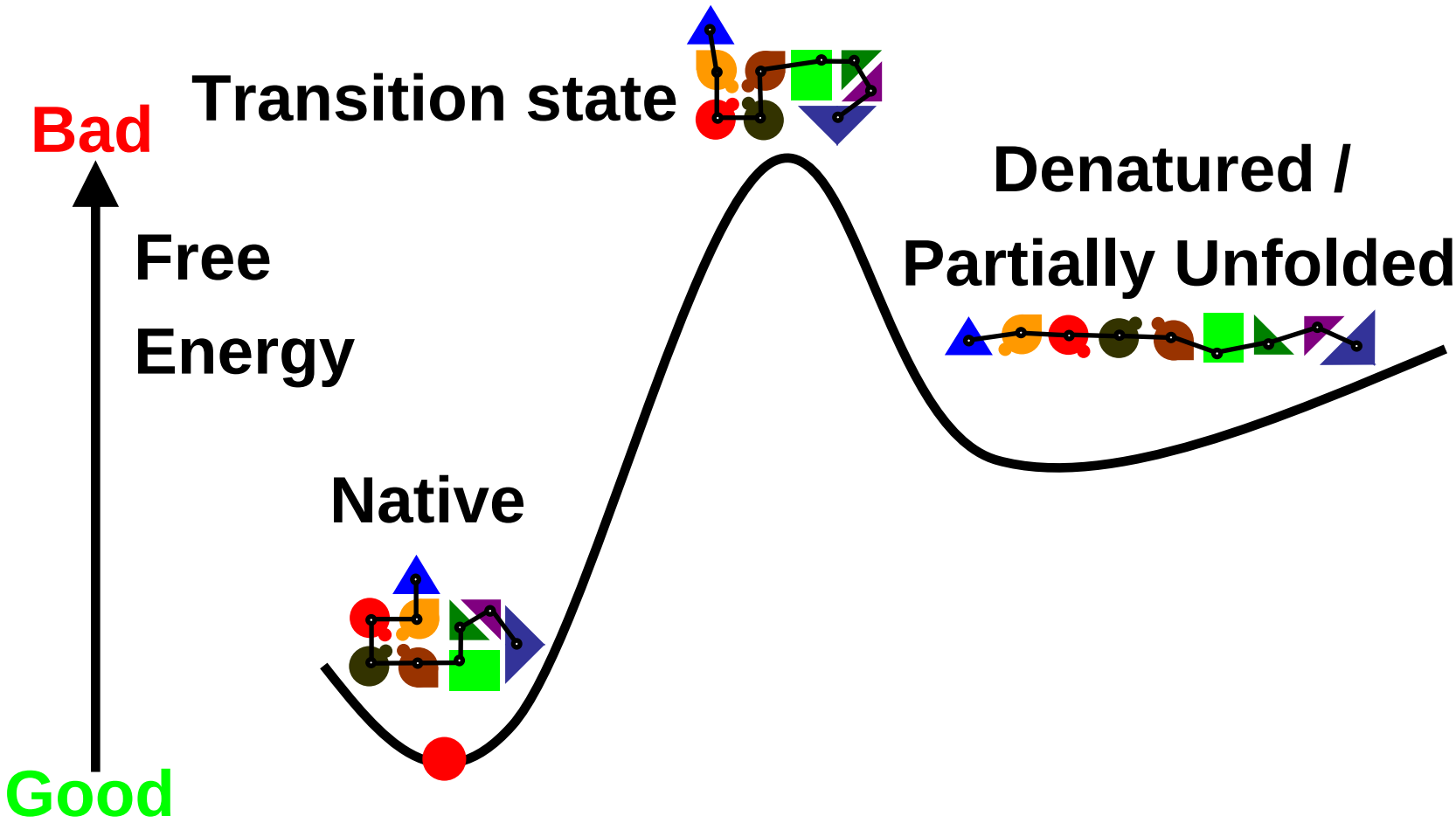
Proteins

- Protein folding as a reaction



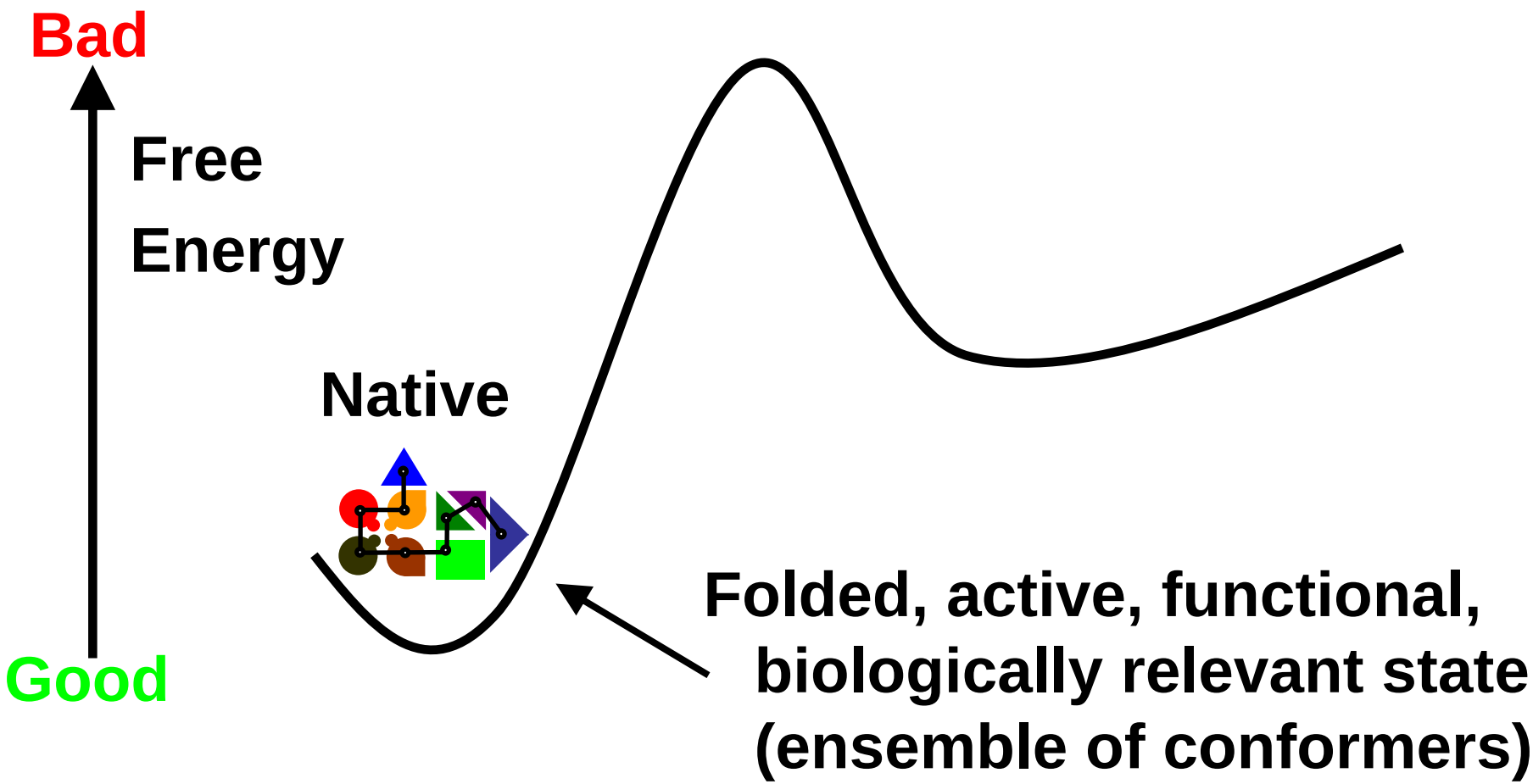
Proteins

- **Protein folding …**



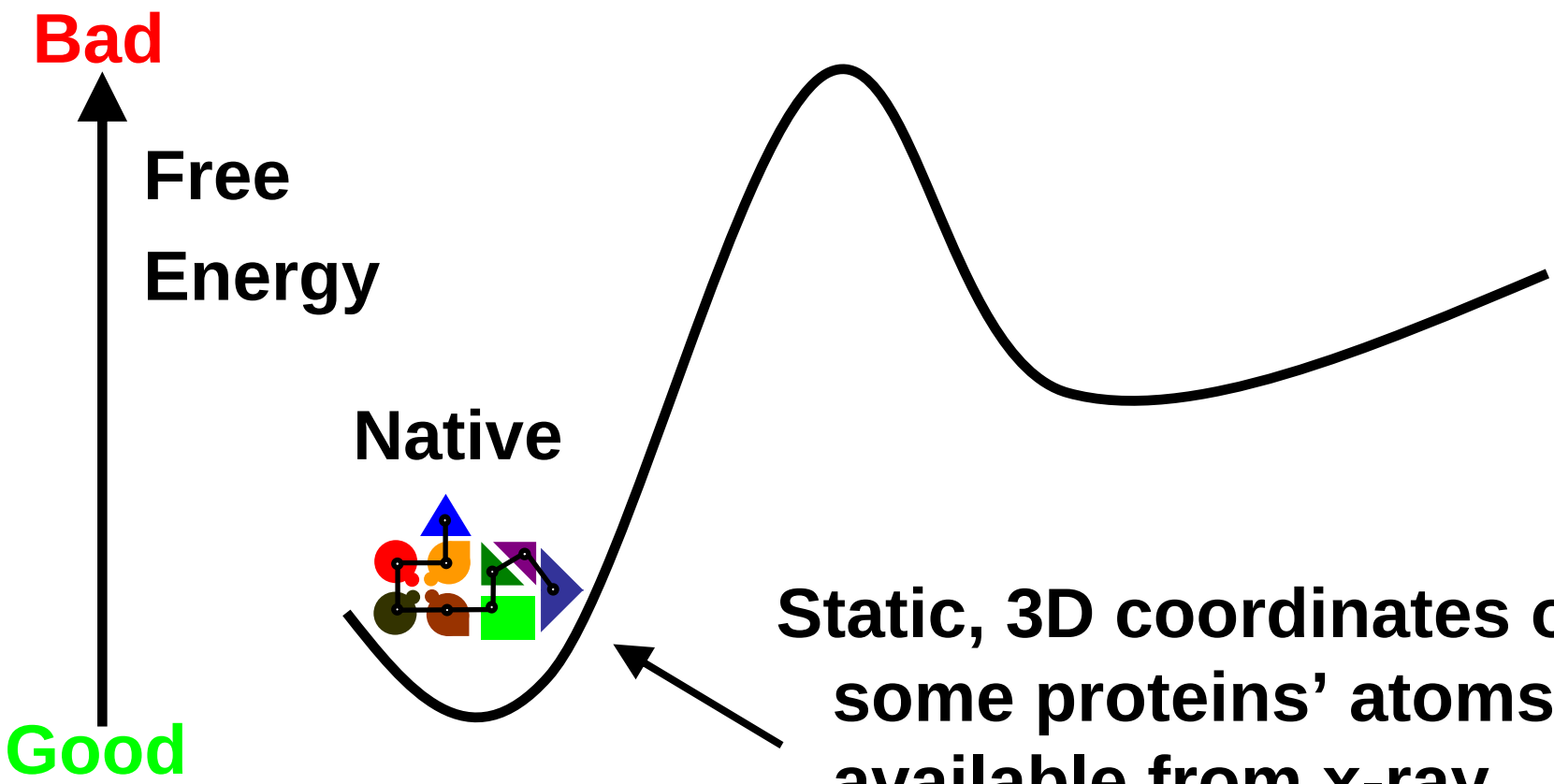
Proteins

- Folded proteins



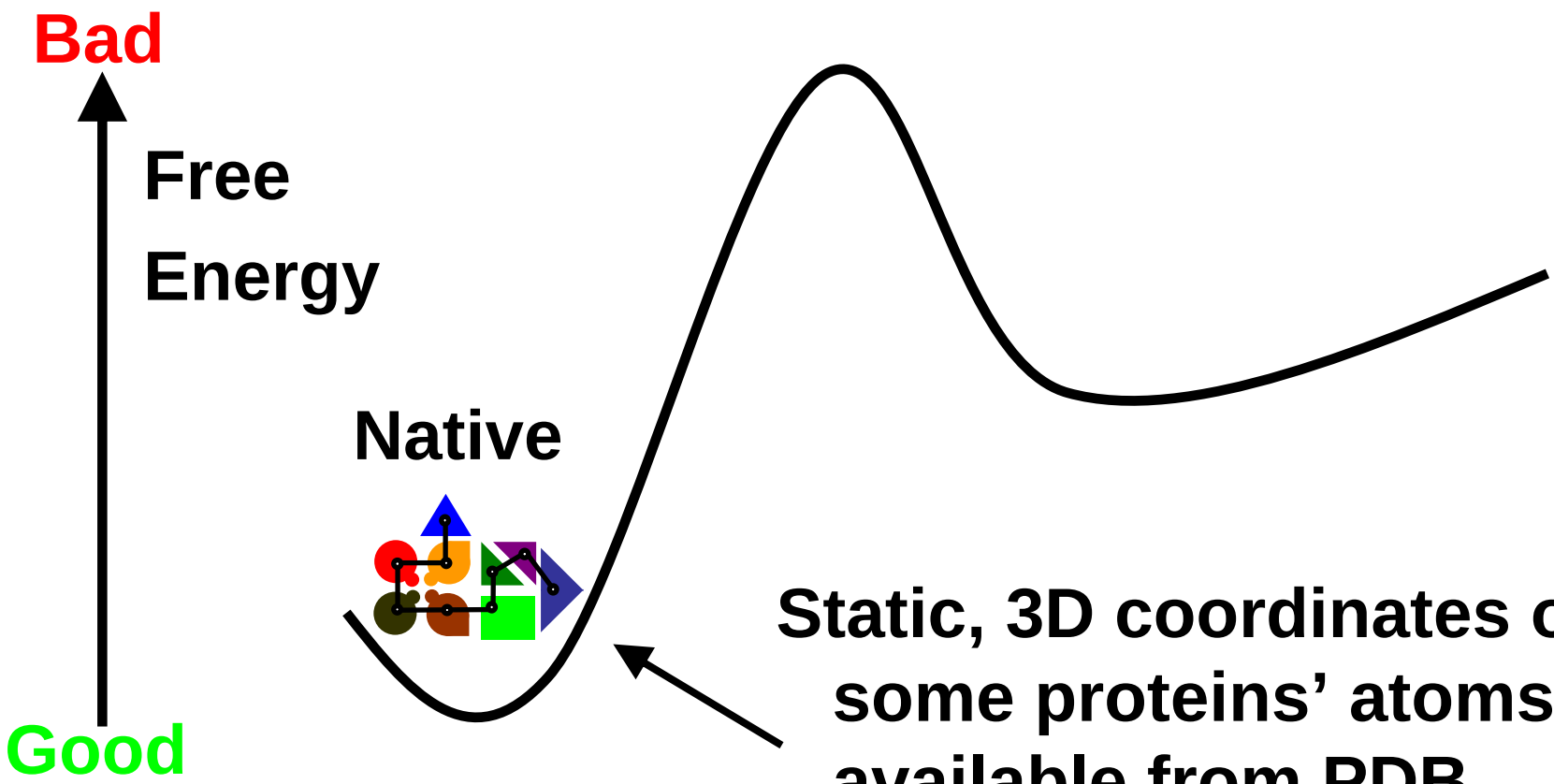
Proteins

- **Folded proteins**



Proteins

- **Folded proteins**

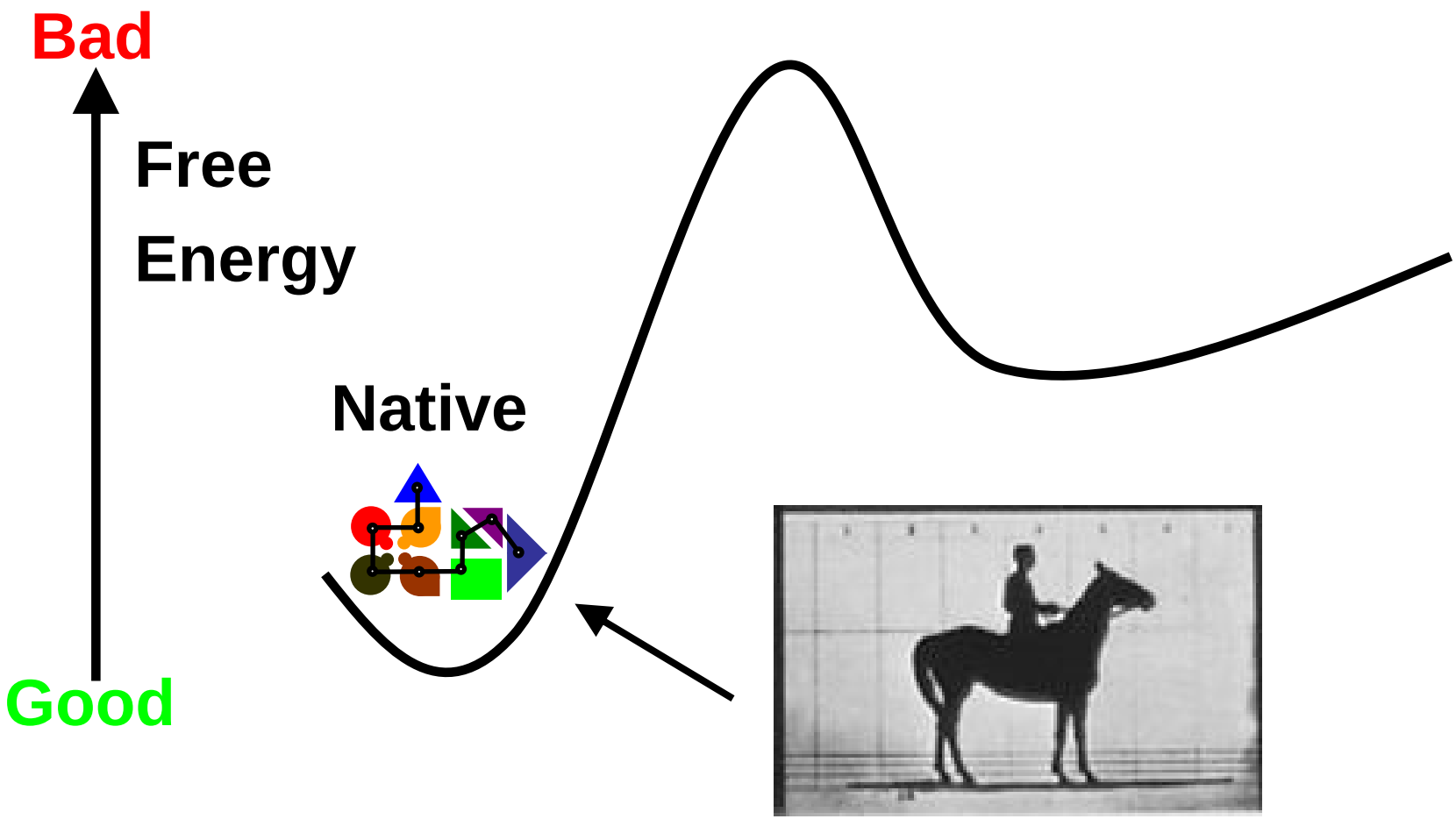


Static, 3D coordinates of some proteins' atoms are available from PDB

http://www.pdb.org

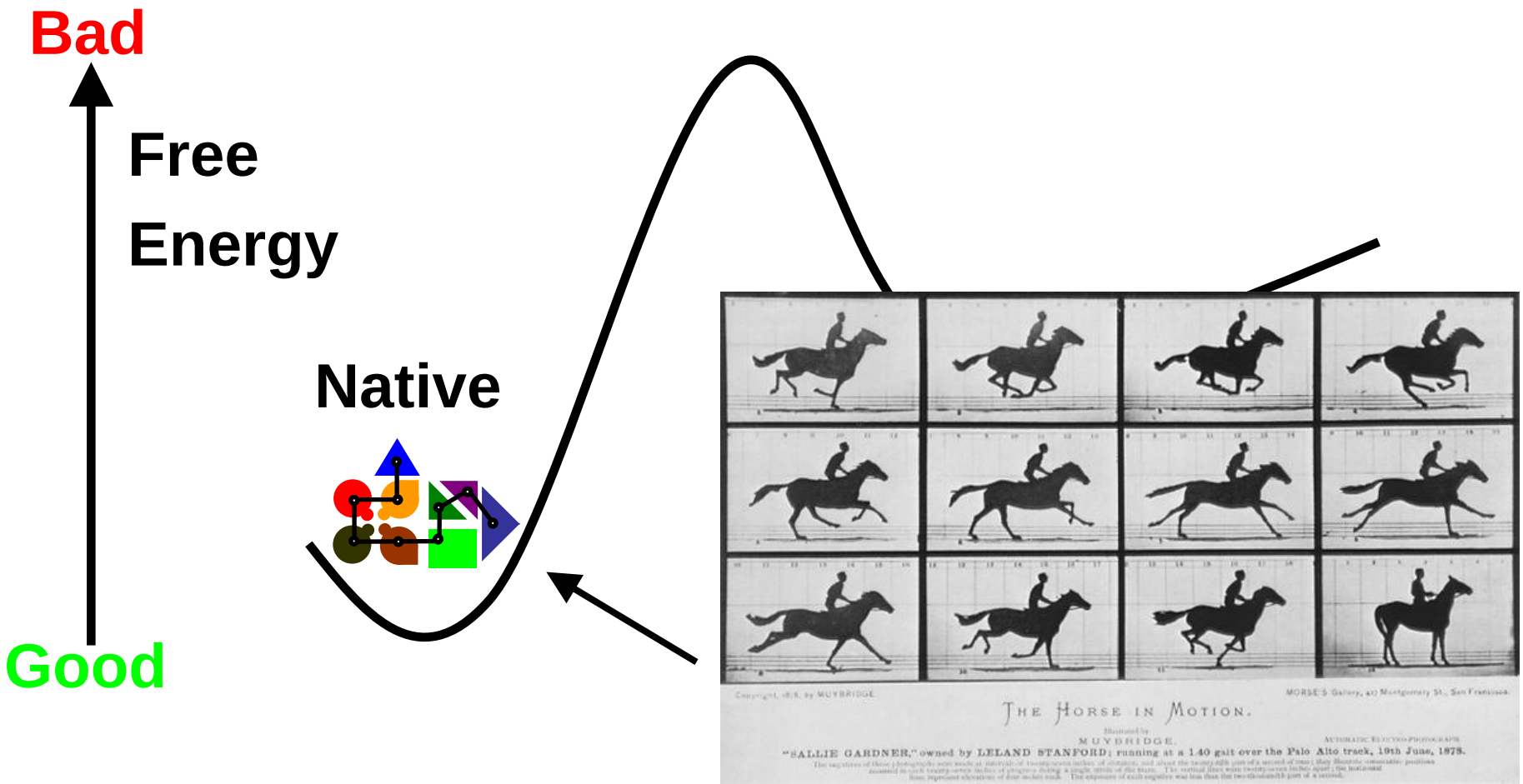
Proteins

- **Folded proteins are complex and dynamic molecules**



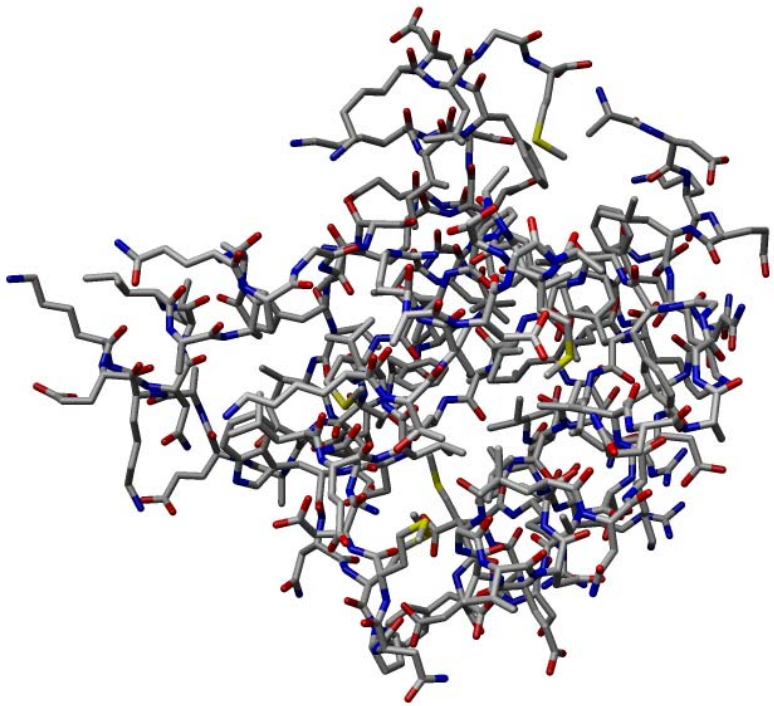
Proteins

- **Folded proteins are complex and dynamic molecules**



Molecular Dynamics

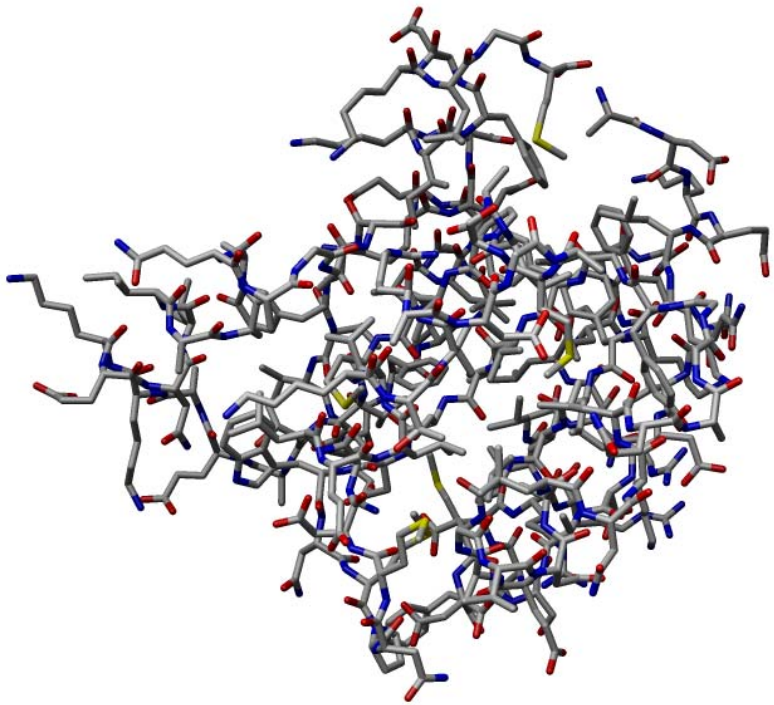
- **MD provides atomic resolution of native dynamics**

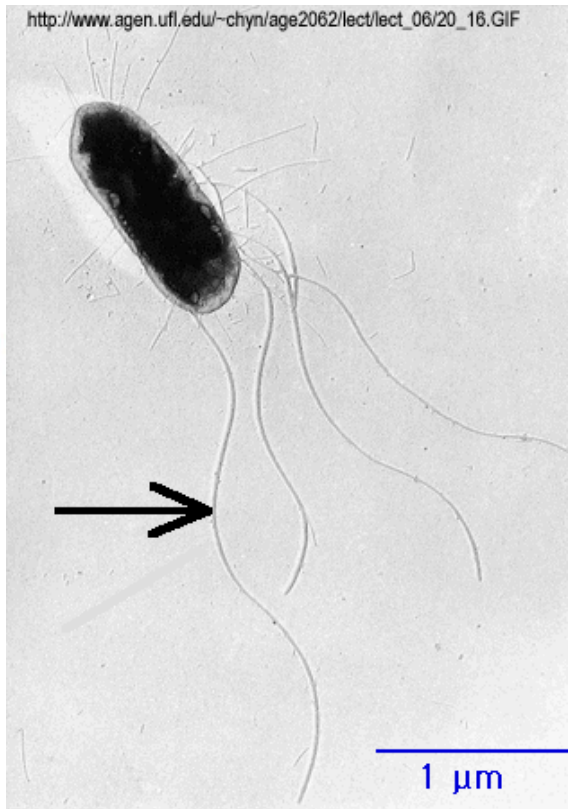


PDB ID: 3chy, *E.* coli CheY 1.66 Å X-ray crystallography

Molecular Dynamics

- **MD provides atomic resolution of native dynamics**

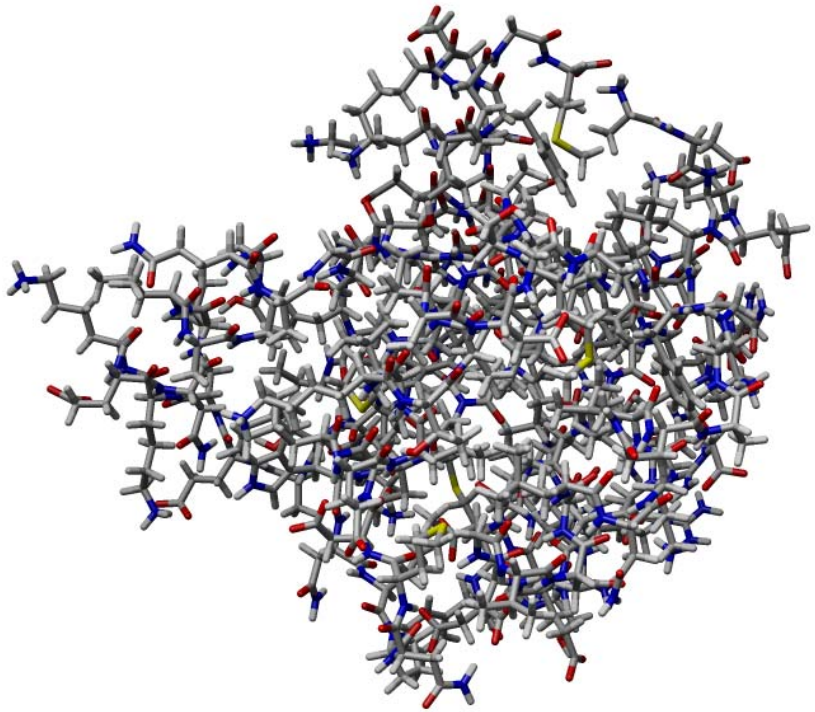




PDB ID: 3chy, *E.* coli CheY 1.66 Å X-ray crystallography

Molecular Dynamics

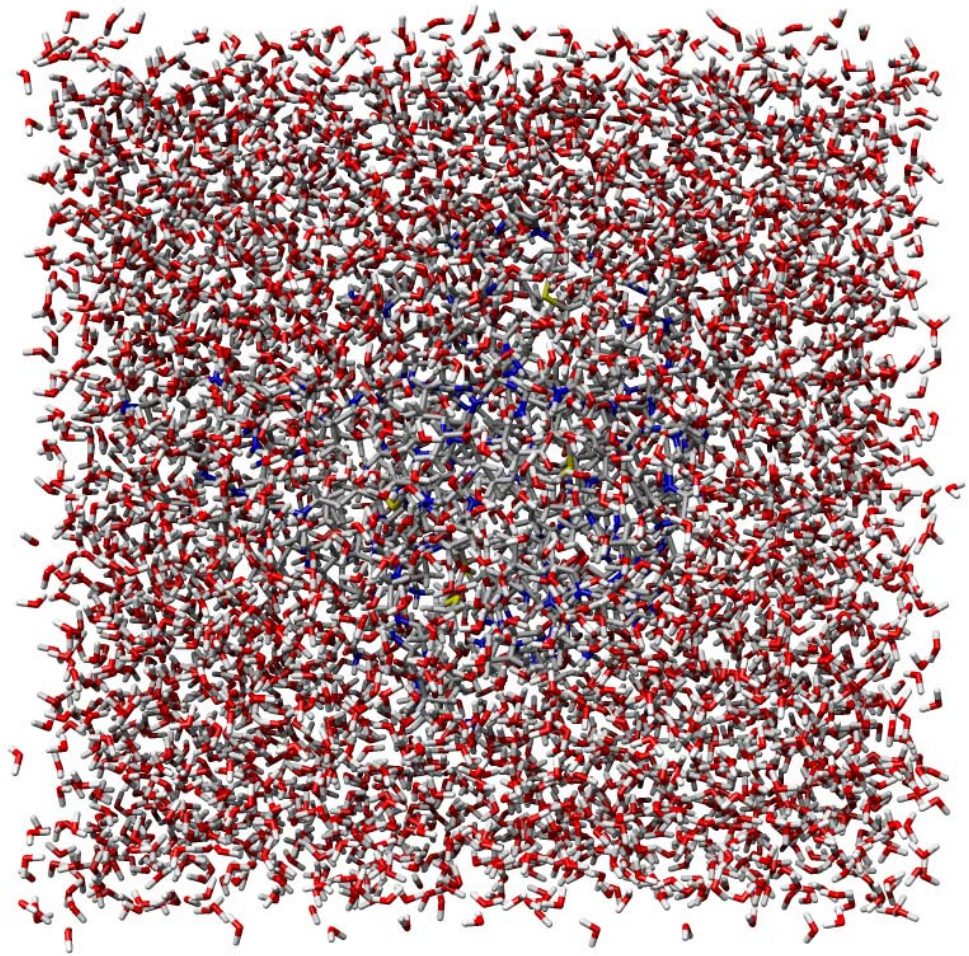
- **MD provides atomic resolution of native dynamics**



3chy, hydrogens added

Molecular Dynamics

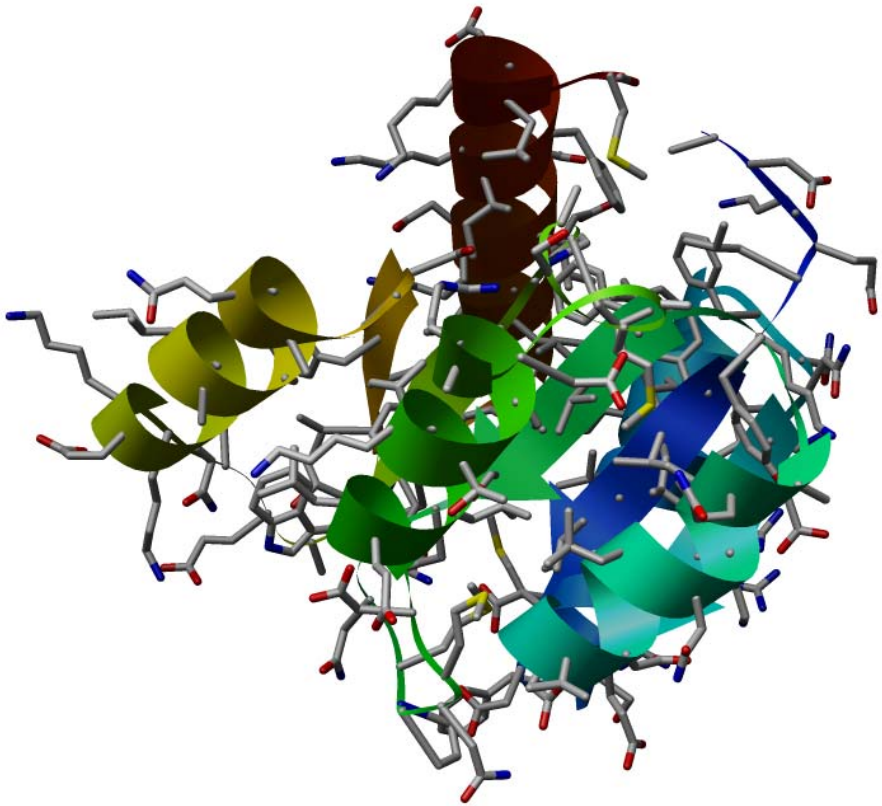
- **MD provides atomic resolution of native dynamics**



3chy, waters added (i.e. solvated)

Molecular Dynamics

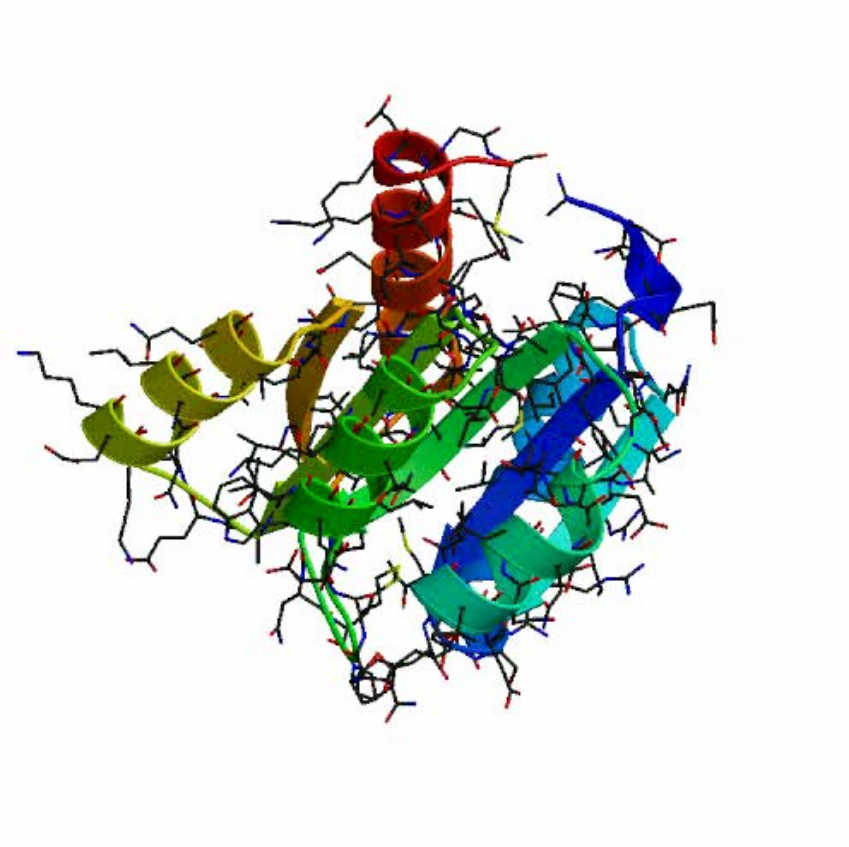
- **MD provides atomic resolution of native dynamics**



3chy, waters and hydrogens hidden

Molecular Dynamics

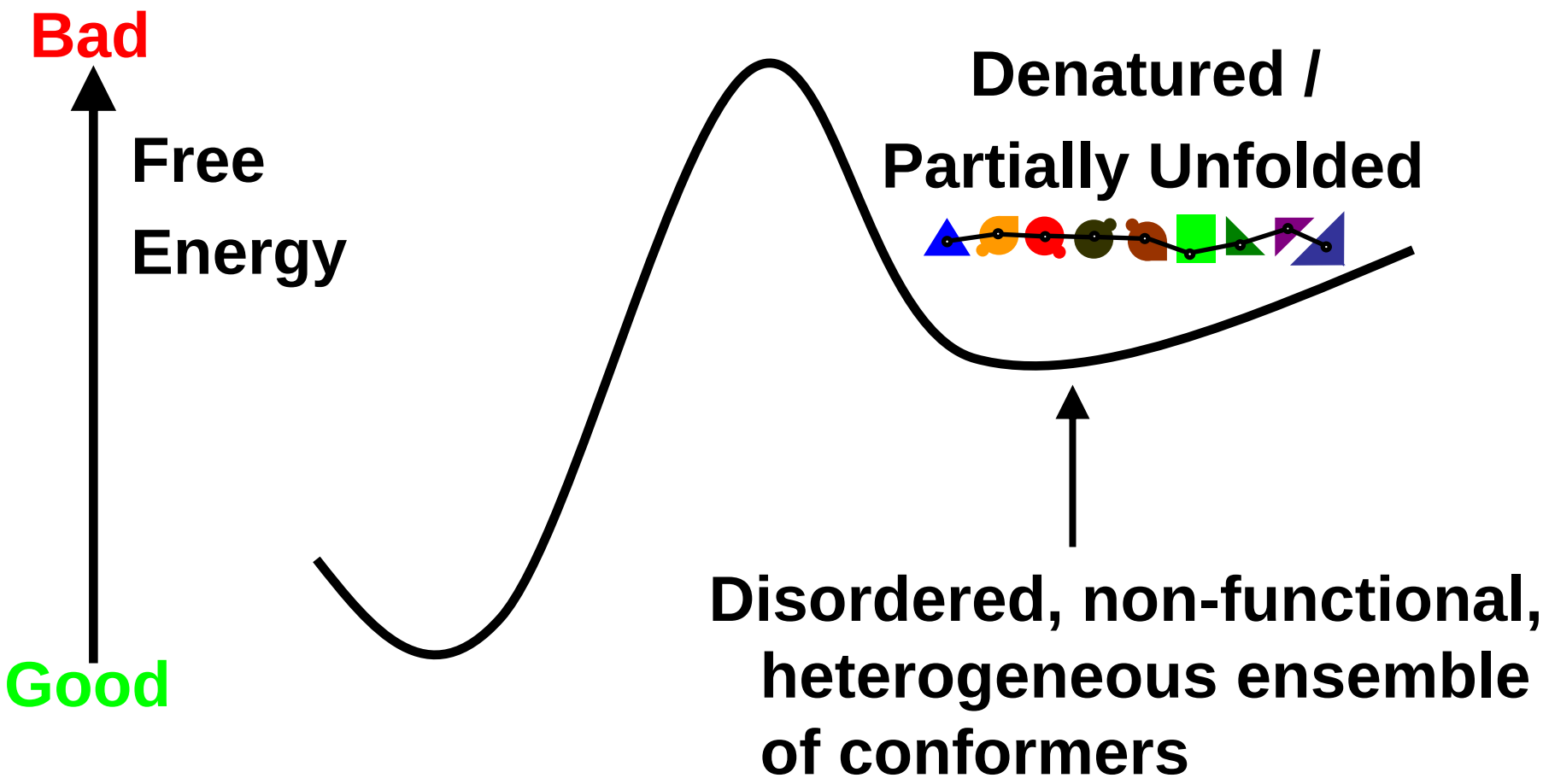
- **MD provides atomic resolution of native dynamics**



native state simulation of 3chy at 298 Kelvin, waters and hydrogens hidden

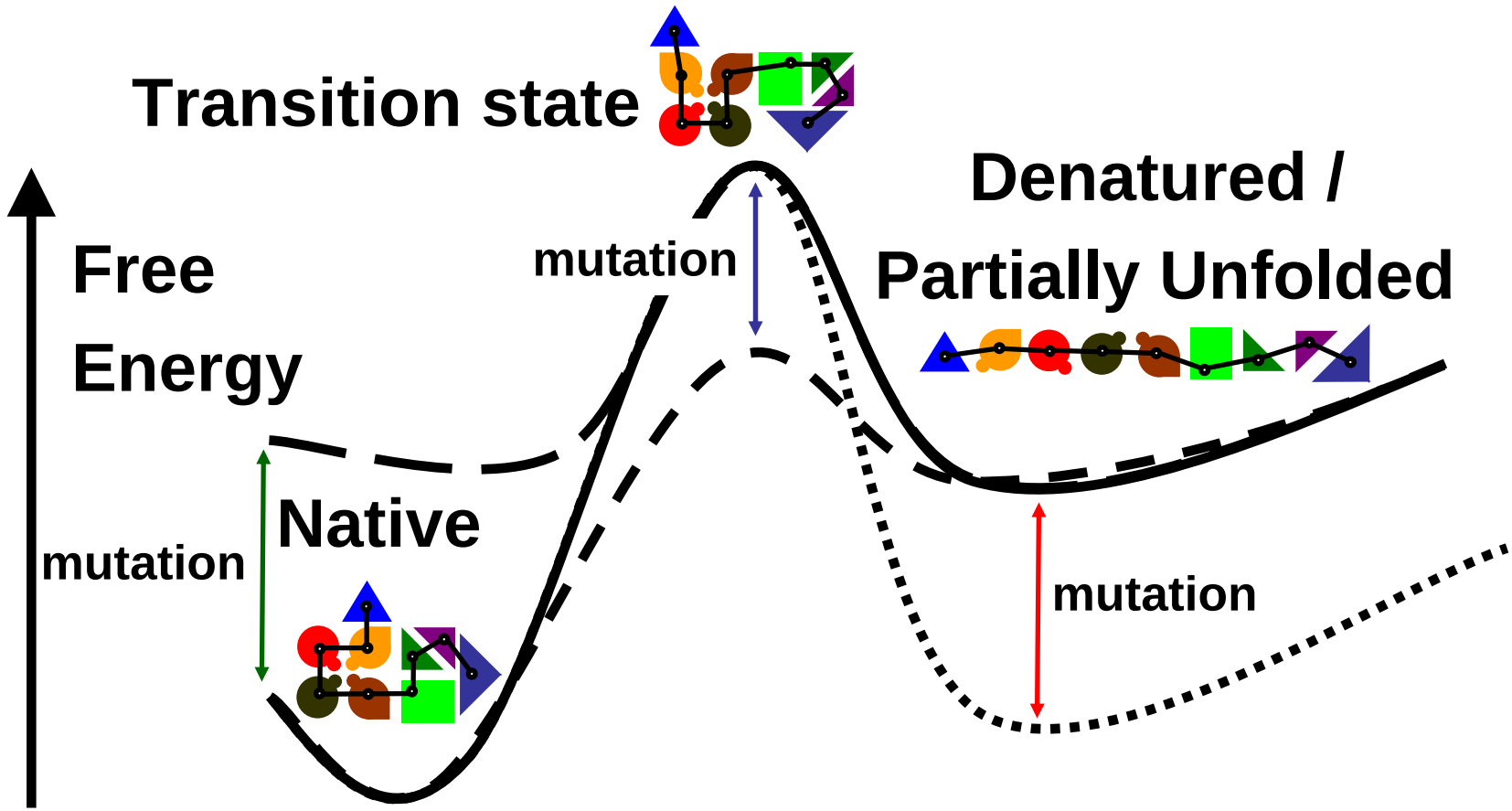
Proteins

- **Folding & unfolding at atomic resolution**



Proteins

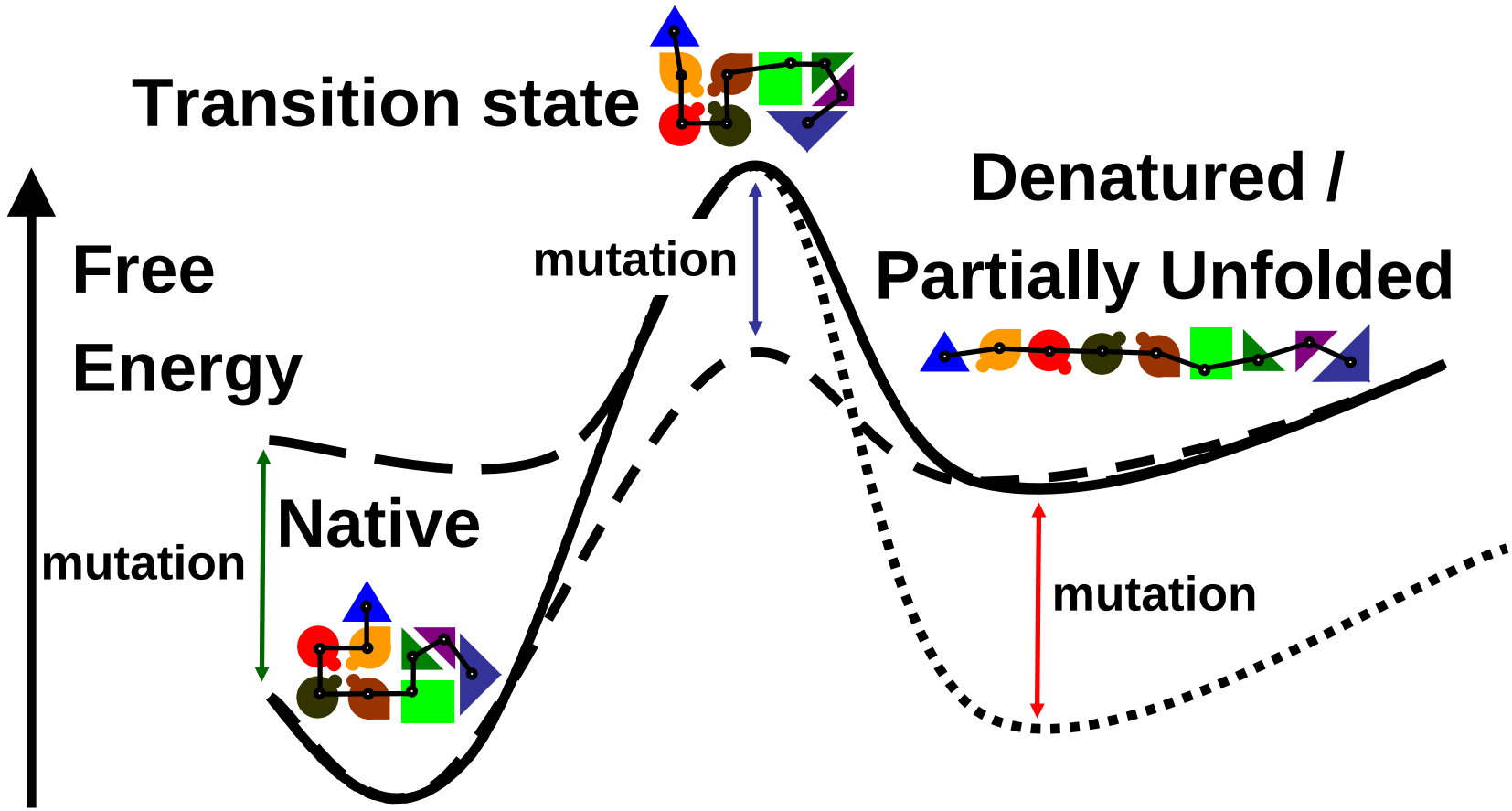
- Protein folding, why we care how it happens



Many diseases are related to protein folding and / or misfolding in response to genetic mutation.

Proteins

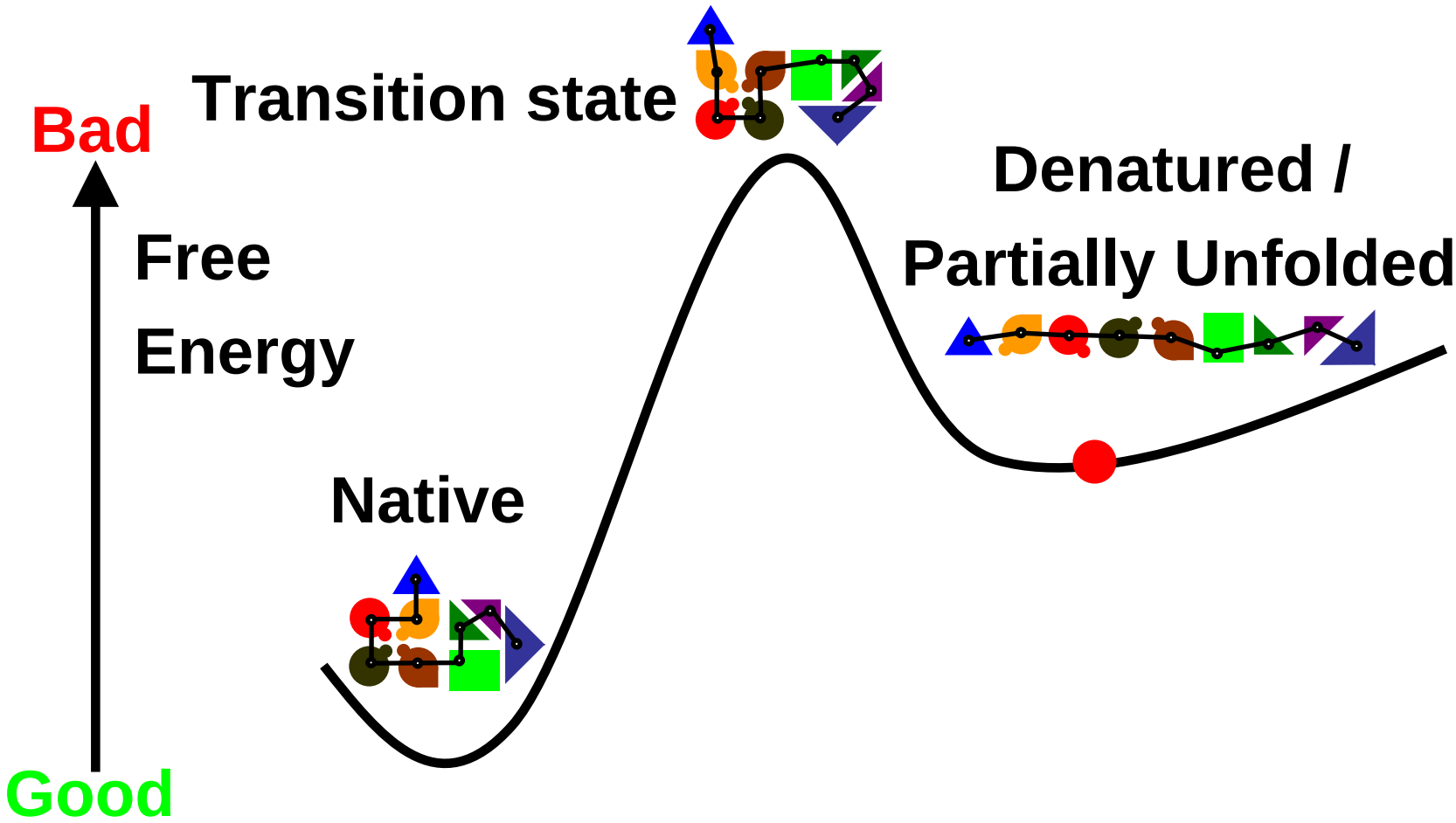
- Protein folding, why we care how it happens



We need to comprehend folding to build nano-scale biomachines (that could produce energy, etc…)

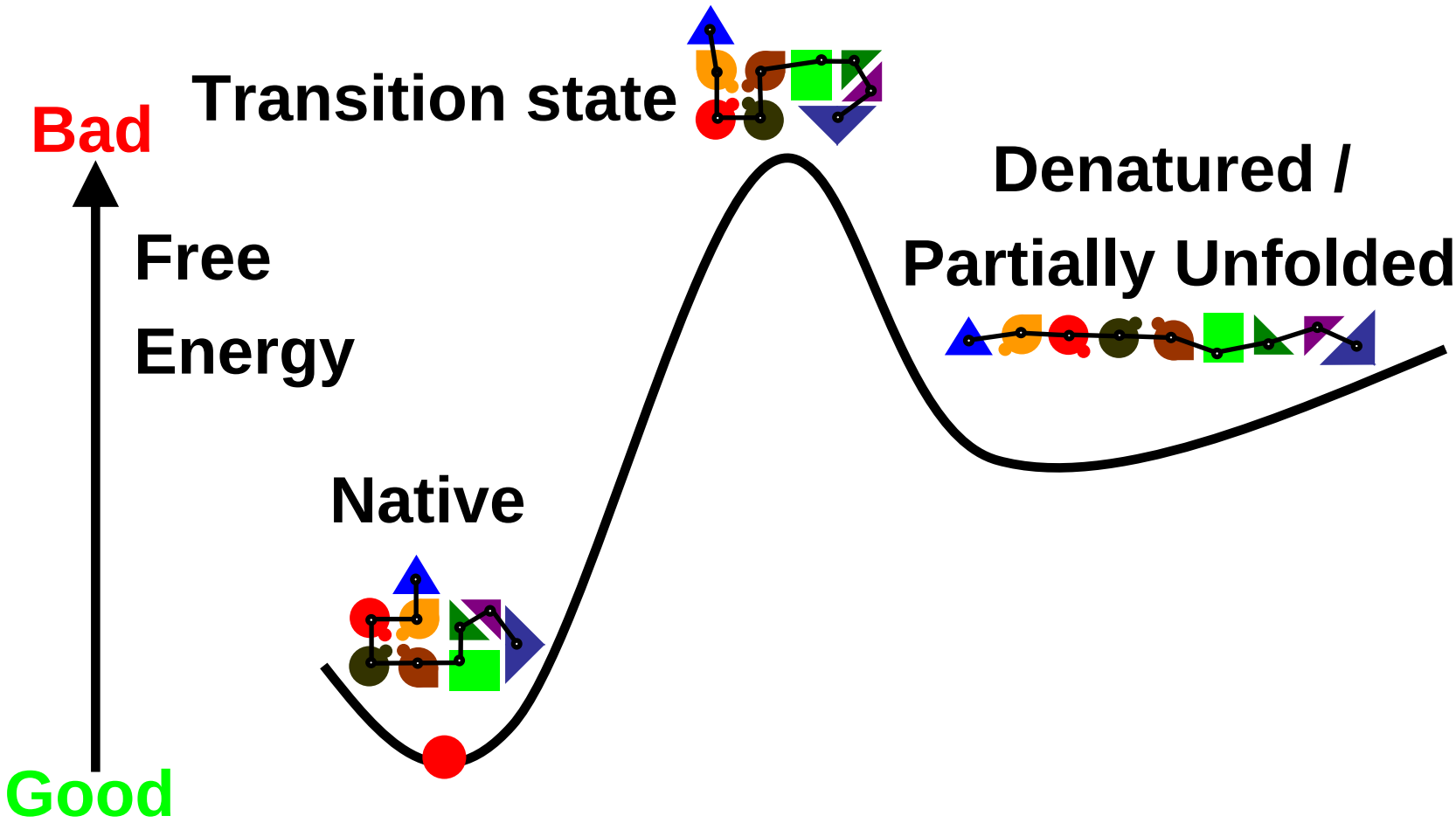
Proteins

- Protein folding takes > 10 μs (often much longer)



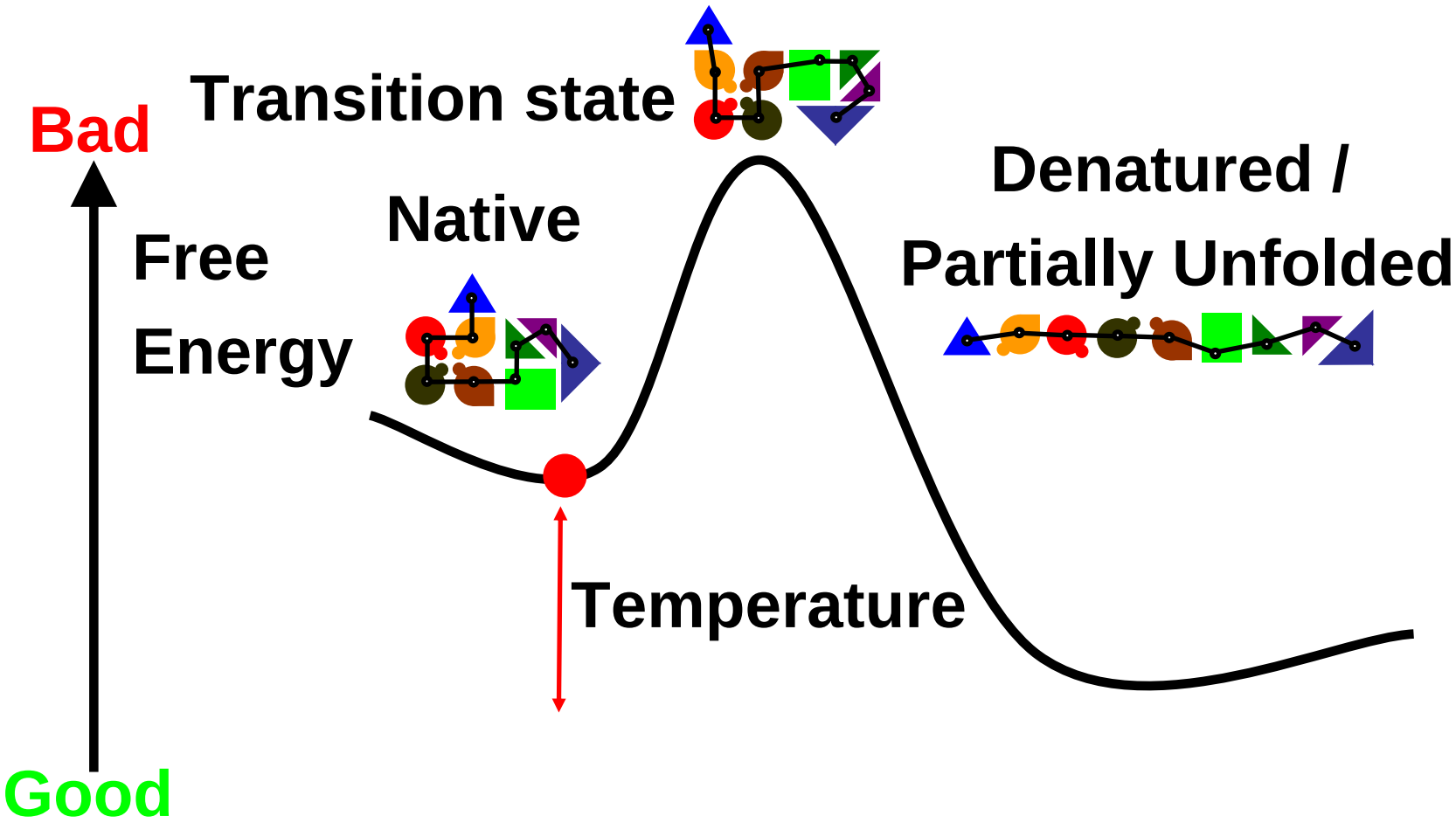
Proteins

- Protein folding is the reverse of protein unfolding



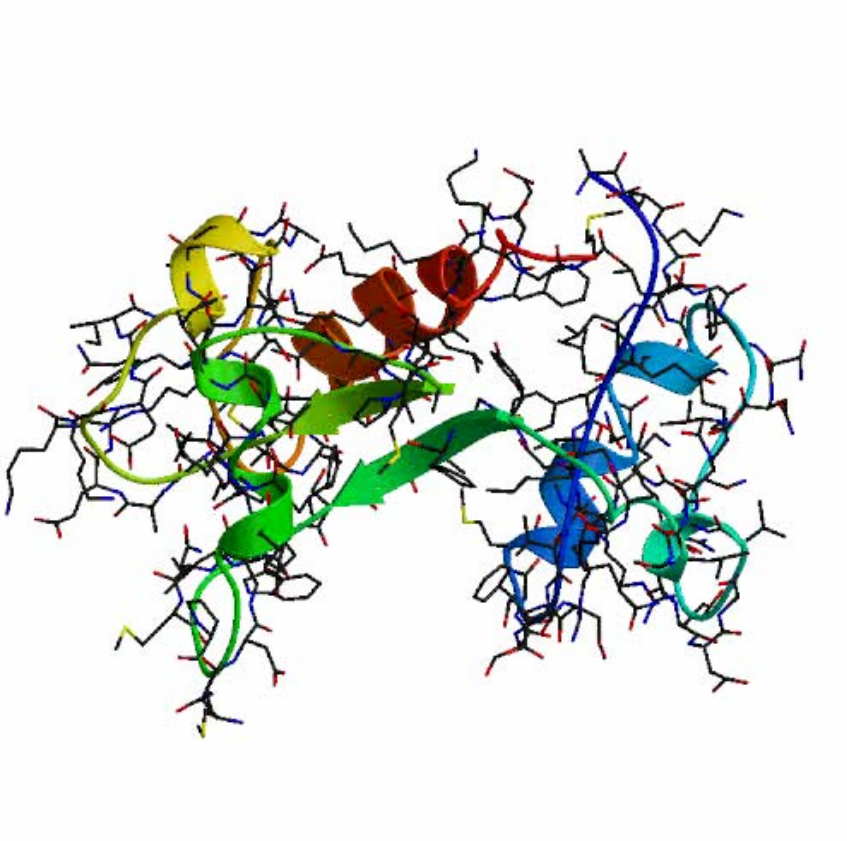
Proteins

- **Protein unfolding is relatively invariant to temperature**



Molecular Dynamics

- **MD provides atomic resolution of folding / unfolding**



unfolding simulation (reversed) of 3chy at 498 Kelvin, waters & hydrogens hidden

Molecular Dynamics[1]

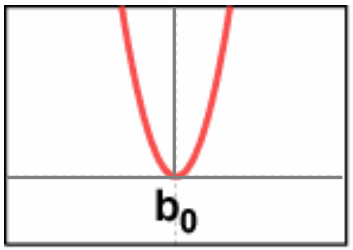
- Classically evolves an atomic system with time
  - **Potential function (a.k.a force field)**
    - Describes the energies of interaction between atom centers
  - **Integration algorithm**
    - Time dependent evolution of atomic coordinates in response to potential energy
  - **Statistical sampling ensemble**
    - Fixed thermodynamic variables, i.e. NVE
    - Number of atoms, box Volume, total Energy

1. Beck, D.A.C. Daggett, V. Methods (2004) 31: 112-120

Molecular Dynamics

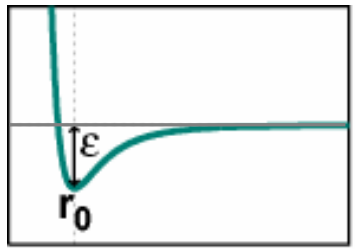
- **Potential function for MD[1,2]**

***U* = Bond + Angle + Dihedral + van der Waals + Electrostatic**



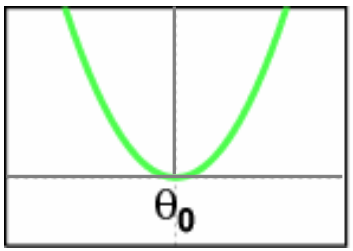
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



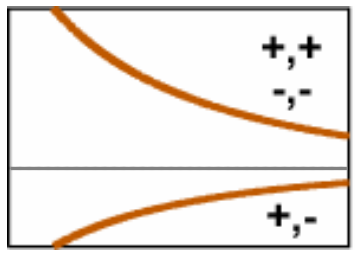
van der Waals

$$\sum_{pairs\cdot i,j} [\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}]$$



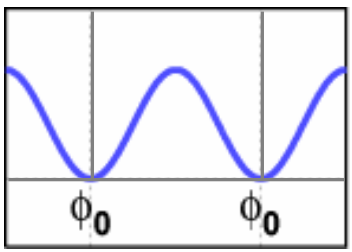
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{\theta,i}(\theta_i - \theta_{0,i})^2$$



Electrostatic

$$332 \sum_{pairs\cdot i,j} \left(\frac{q_i q_j}{r_{ij}}\right)$$



Dihedral

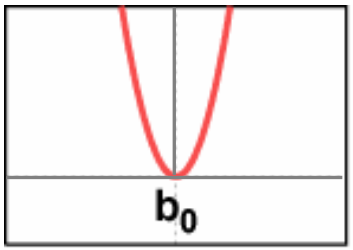
$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91: 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101: 5051-5061

Molecular Dynamics

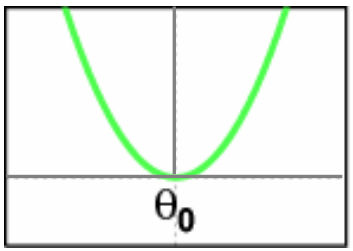
- **Potential function for MD[1,2]**

***U* = Bond + Angle + Dihedral + van der Waals + Electrostatic**



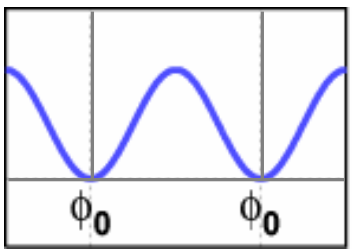
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



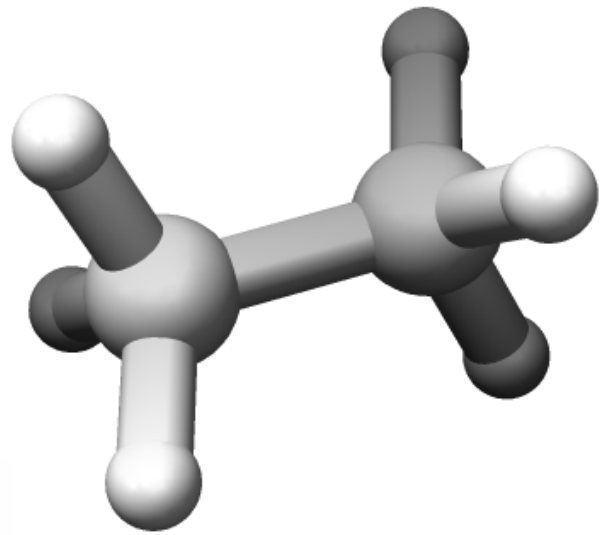
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{\theta,i}(\theta_i - \theta_{0,i})^2$$



Dihedral

$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

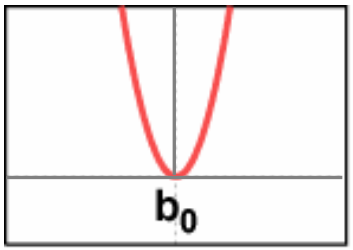


1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91: 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101: 5051-5061

Molecular Dynamics

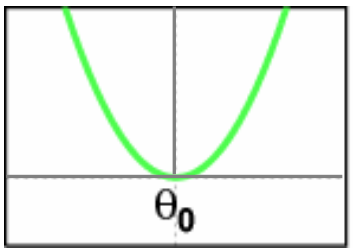
- **Potential function for MD[1,2]**

***U* = Bond + Angle + Dihedral + van der Waals + Electrostatic**



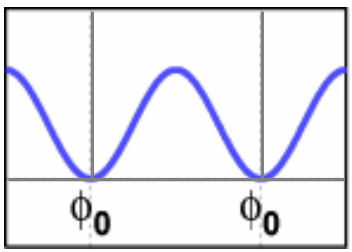
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



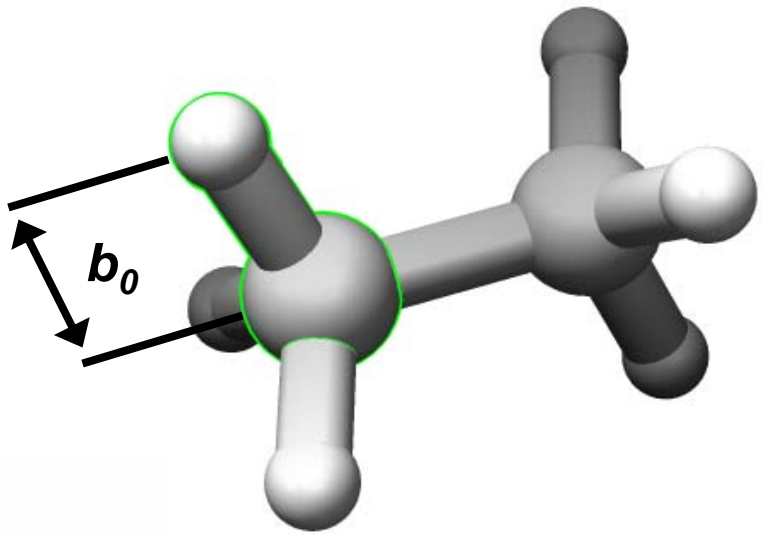
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{\theta,i}(\theta_i - \theta_{0,i})^2$$



Dihedral

$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

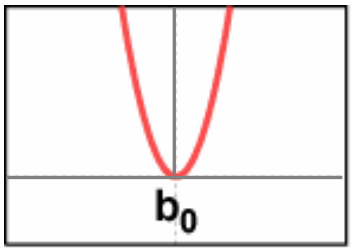


1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91: 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101: 5051-5061

Molecular Dynamics

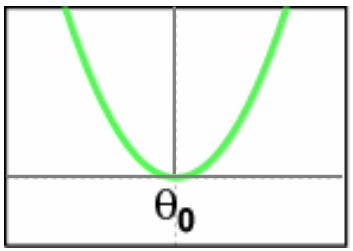
- **Potential function for MD[1,2]**

***U* = Bond + Angle + Dihedral + van der Waals + Electrostatic**



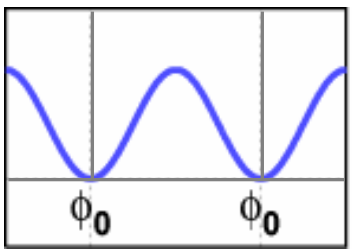
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



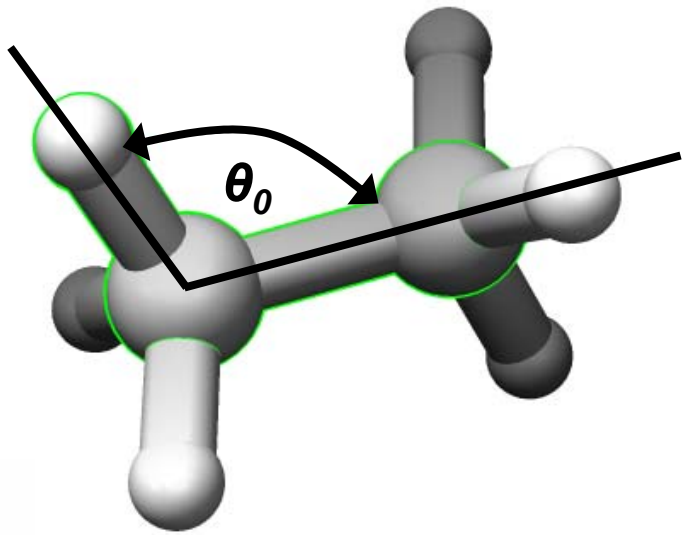
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{\theta,i}(\theta_i - \theta_{0,i})^2$$



Dihedral

$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

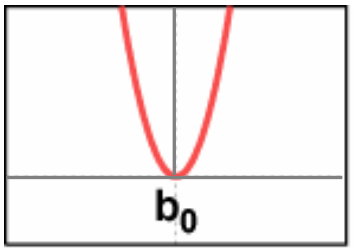


1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91: 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101: 5051-5061

Molecular Dynamics

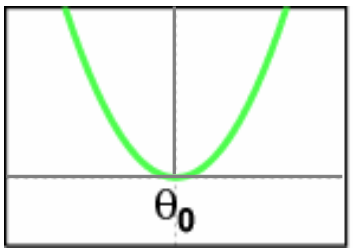
- **Potential function for MD[1,2]**

***U* = Bond + Angle + Dihedral + van der Waals + Electrostatic**



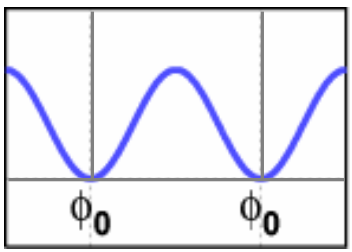
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



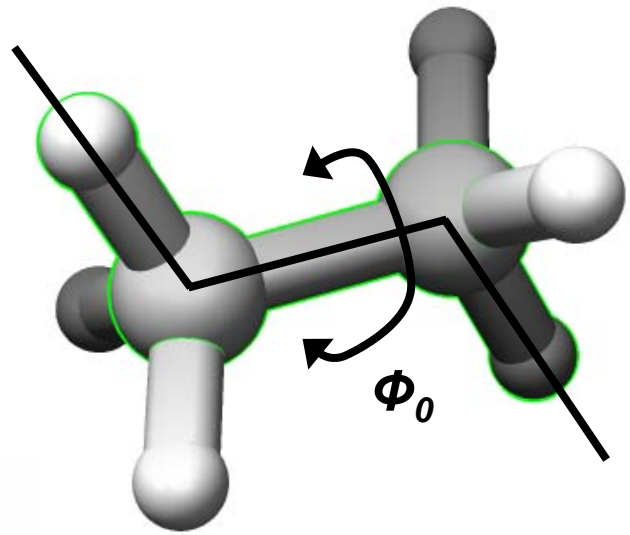
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{,i}(\theta_i - \theta_{0,i})^2$$



Dihedral

$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

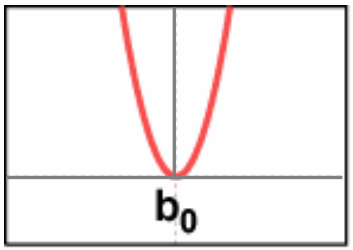


1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101:25 5051-5061

Molecular Dynamics

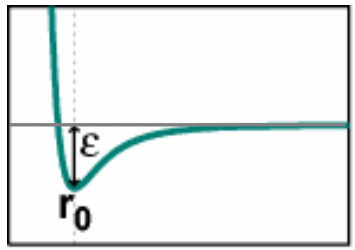
- **Potential function for MD[1,2]**

U **= Bond + Angle + Dihedral +** **van der Waals + Electrostatic**



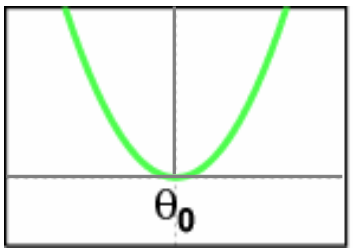
Bond

$$\sum_{i}^{bonds} K_{b,i}(b_i - b_{0,i})^2$$



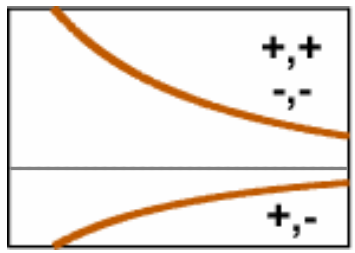
van der Waals

$$\sum_{pairs\cdot i,j} [\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}]$$



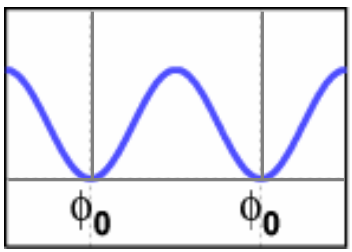
Angle

$$\sum_{i}^{\substack{bond \\ angles}} K_{,i}(\theta_i - \theta_{0,i})^2$$



Electrostatic

$$332 \sum_{pairs\cdot i,j} \left(\frac{q_i q_j}{r_{ij}}\right)$$



Dihedral

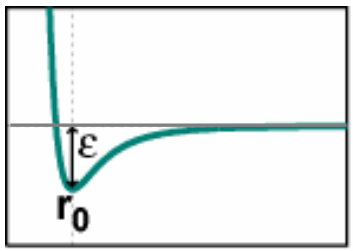
$$\sum_{i}^{\substack{torsion \\ angles}} K_{\phi,i}\{1 - \cos[n_i(\phi_i - \phi_{0,i})]\}$$

1. Levitt M. Hirshberg M. Sharon R. Daggett V. Comp. Phys. Comm. (1995) 91: 215-231
2. Levitt M. *et al.* J. Phys. Chem. B (1997) 101: 5051-5061

Molecular Dynamics

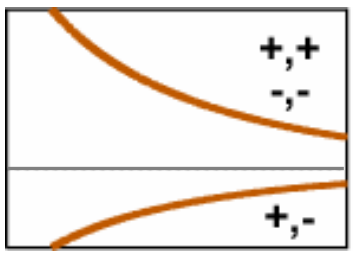
- **Non-bonded components of potential function**

  U_{nb} = van der Waals + Electrostatic



van der Waals

$$\sum_{pairs\cdot i,j} [\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}]$$



Electrostatic

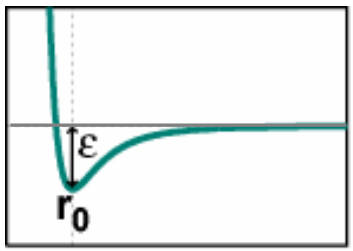
$$332 \sum_{pairs\cdot i,j} \left(\frac{q_i q_j}{r_{ij}}\right)$$

- ***To a large degree, protein structure is dependent on non-bonded atomic interactions***

Molecular Dynamics

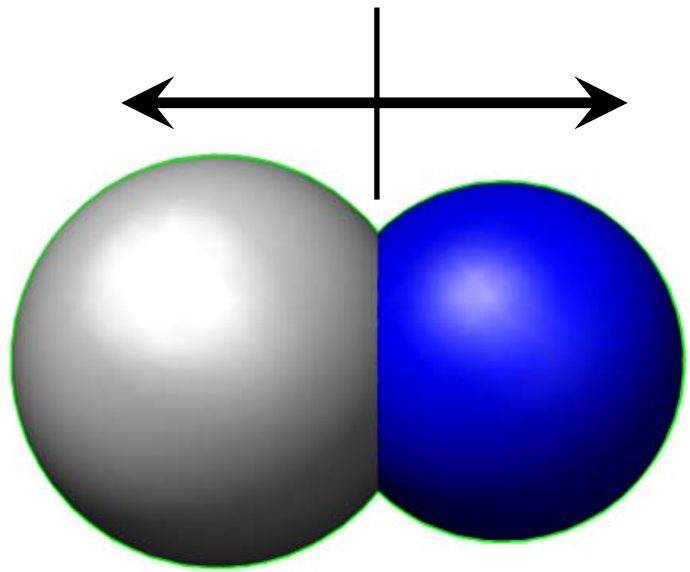
- **Non-bonded components of potential function**

  ***U_{nb}* = van der Waals + Electrostatic**



van der Waals

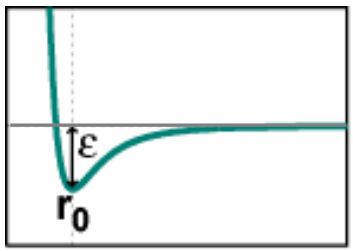
$$\sum_{pairs\text{-}i,j} [\varepsilon_{ij} \left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2\varepsilon_{ij} \left(\frac{r_{0,ij}}{r_{ij}} \right)^{6}]$$



Molecular Dynamics

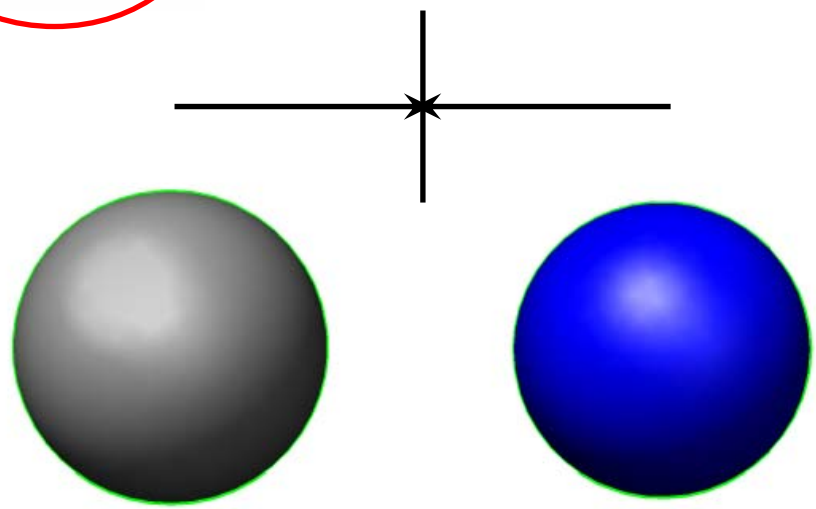
- **Non-bonded components of potential function**

  ***U_{nb}* = van der Waals + Electrostatic**



van der Waals

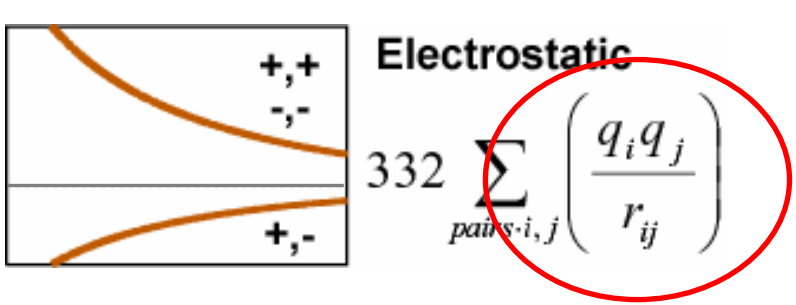
$$\sum_{pairs\text{-}i,j} [\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}]$$

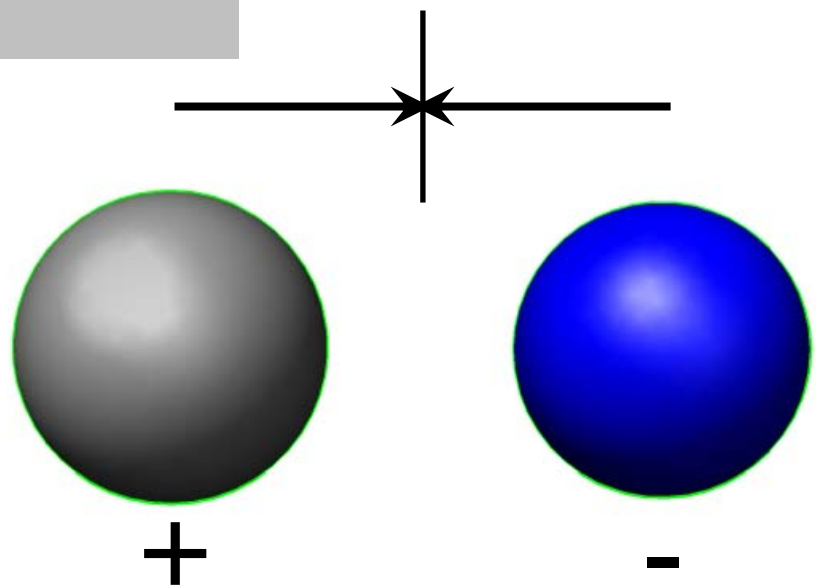


Molecular Dynamics

- **Non-bonded components of potential function**

  U_{nb} **= van der Waals + Electrostatic**

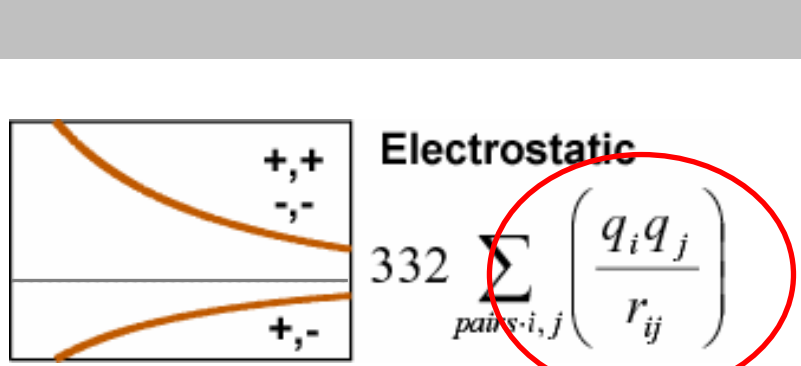




Molecular Dynamics

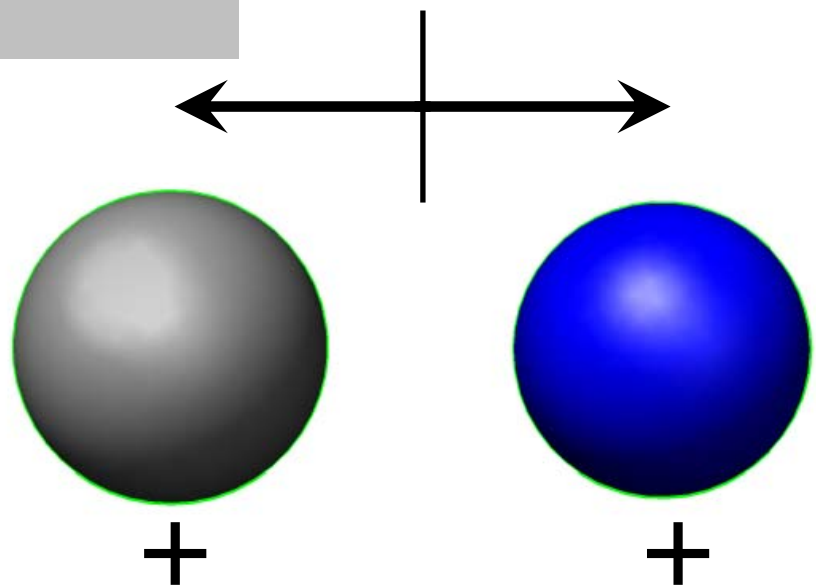
- **Non-bonded components of potential function**

  U_{nb} **= van der Waals + Electrostatic**



Electrostatic

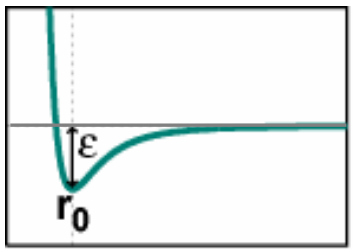
$$332 \sum_{pairs\ i,j} \left(\frac{q_i q_j}{r_{ij}} \right)$$



Molecular Dynamics

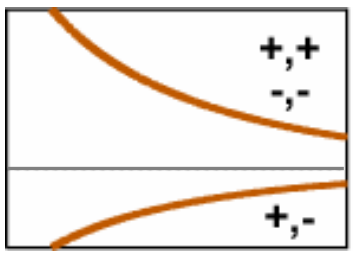
- **Non-bonded components of potential function**

U_{nb} = van der Waals + Electrostatic



van der Waals

$$\sum_{pairs\cdot i,j} [\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\varepsilon_{ij}\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}]$$



Electrostatic

$$332 \sum_{pairs\cdot i,j} \left(\frac{q_i q_j}{r_{ij}}\right)$$

NOTE:

Sum over all pairs of N atoms, or

$$\frac{N * N - 1}{2}$$ **pairs**

N is often between $5x10^5$ to $5x10^6$

For $5x10^5$ that is $1.25x10^{11}$ pairs

THAT IS A LOT OF POSSIBLE PAIRS!

Molecular Dynamics

- **Time dependent integration of classical equations of motion**

$$F = -\frac{\partial U}{\partial x}$$

$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$

Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

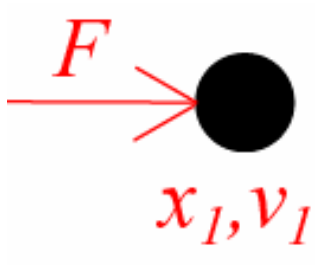
$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

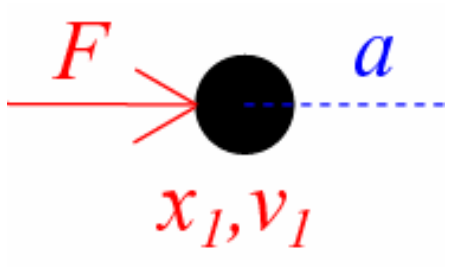
$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

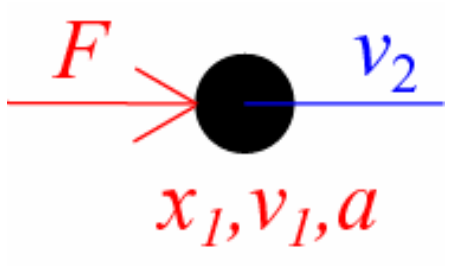
$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

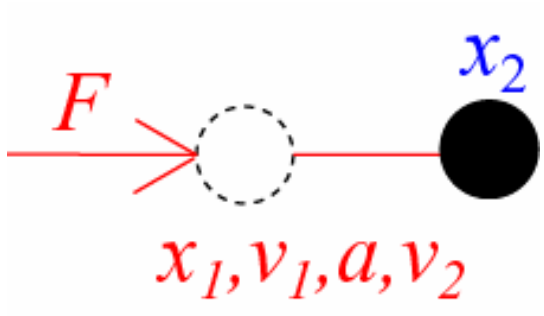
$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

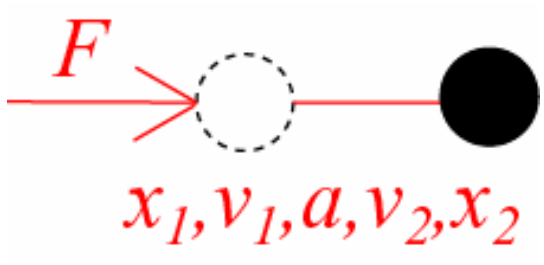
$$F = ma$$

$$a = \frac{v_2 - v_1}{\partial t}$$

$$v = \frac{x_2 - x_1}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

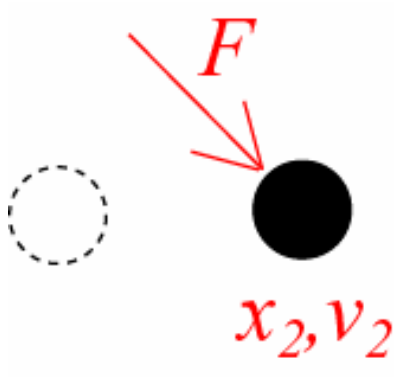
$$F = ma$$

$$a = \frac{v_3 - v_2}{\partial t}$$

$$v = \frac{x_3 - x_2}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$



Molecular Dynamics

- **Time dependent integration**

$$F = -\frac{\partial U}{\partial x}$$

$$F = ma$$

$$a = \frac{v_3 - v_2}{\partial t}$$

$$v = \frac{x_3 - x_2}{\partial t}$$

$$E = U + K$$

$$\partial t = 2 \text{ fs}$$

Evaluate forces and perform integration for every atom

Each picosecond of simulation time requires 500 iterations of cycle

E.g. w/ 50,000 atoms, each ps (10^{-12} s) involves 25,000,000 evaluations

Molecular Dynamics

- **Scalable, parallel MD & analysis software:**

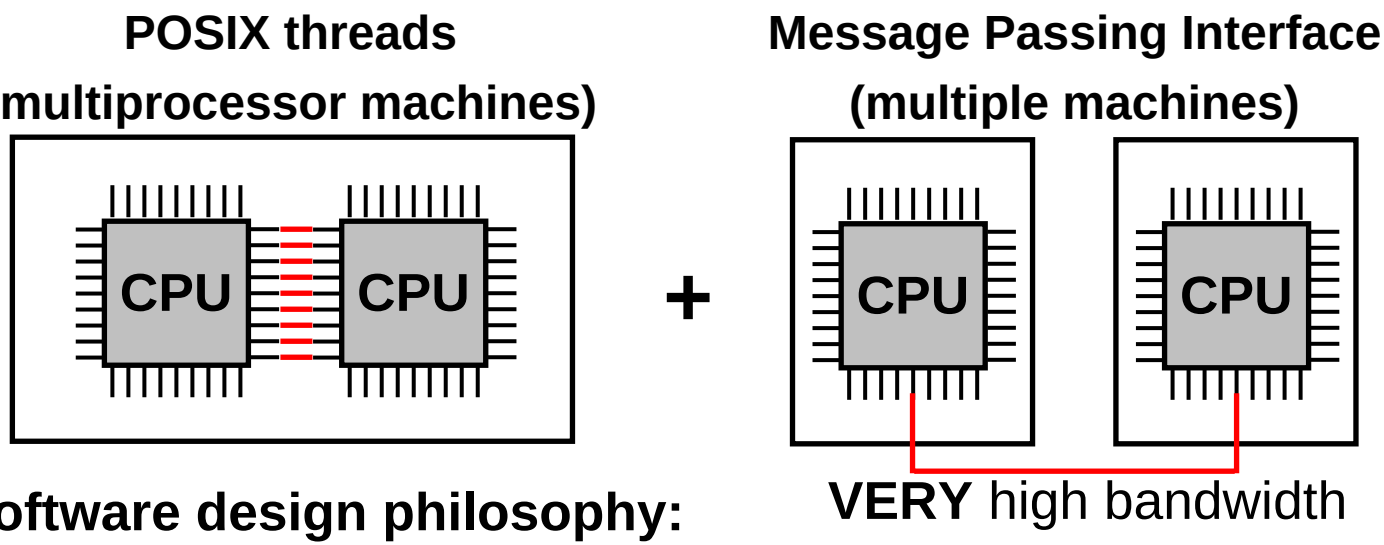
ilmm

in lucem Molecular Mechanics[1]

1. Beck, Alonso, Daggett, (2004) University of Washington, Seattle

Molecular Dynamics

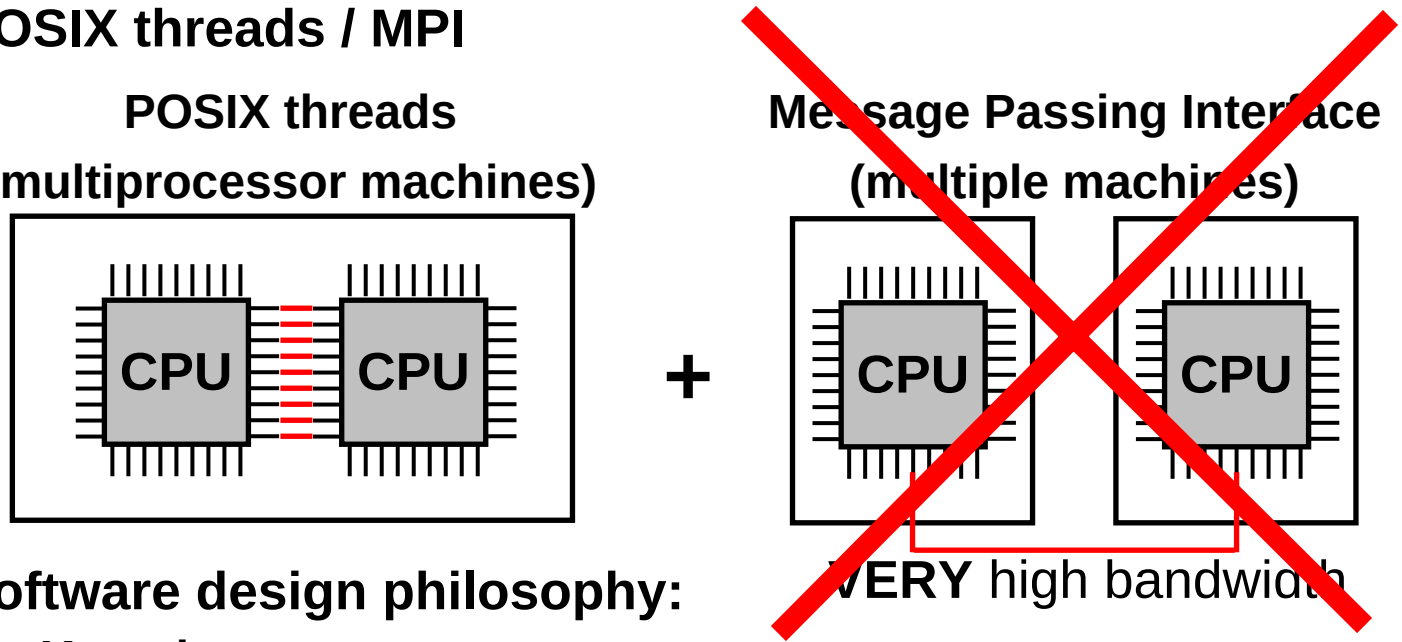
- ***il*mm is written in C (ANSI / POSIX)**
- **64 bit math**
- **POSIX threads / MPI**



- **Software design philosophy:**
  - **Kernel**
    - **Compiles user's molecular mechanics programs**
    - **Schedules execution across processor and machines**
  - **Modules, e.g.**
    - **Molecular Dynamics**
    - **Analysis**

Molecular Dynamics

- ***il*mm is written in C (ANSI / POSIX)**
- **64 bit math**
- **POSIX threads / MPI**



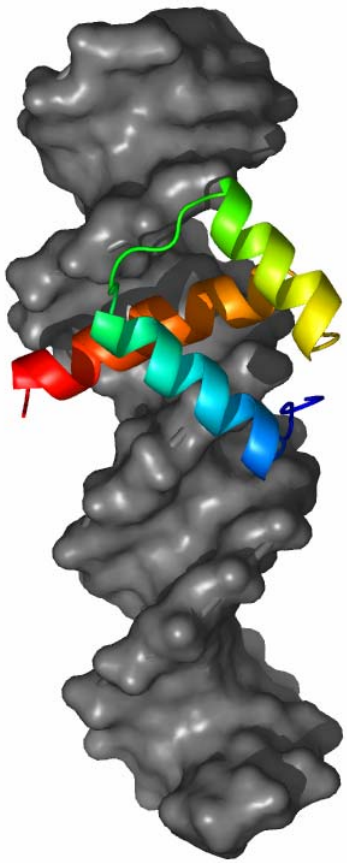
- **Software design philosophy:**
  - **Kernel**
    - **Compiles user's molecular mechanics programs**
    - **Schedules execution across processor and machines**
  - **Modules, e.g.**
    - **Molecular Dynamics**
    - **Analysis**

Dynameomics

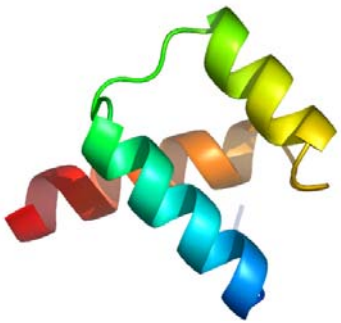
- **Simulate representative protein from all folds**

Dynameomics

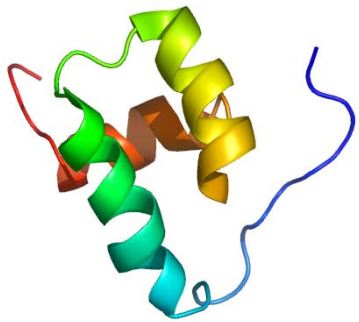
- **Simulate representative protein from all folds**
  - **Nature reuses designs for similar jobs**



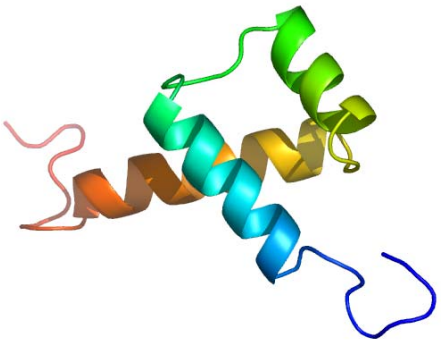
1hdd



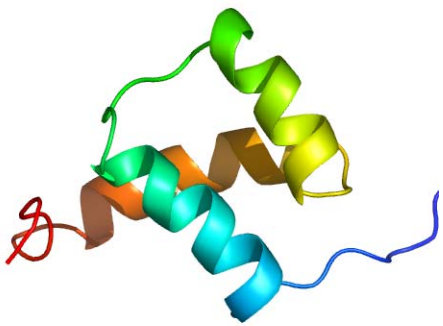
1enh



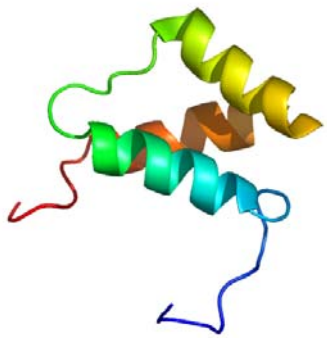
1f43



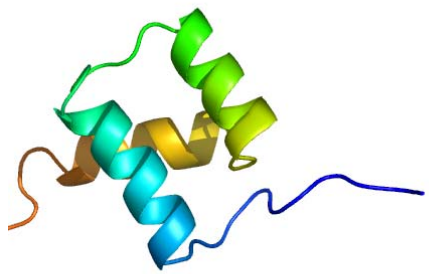
1ftt



1bw5



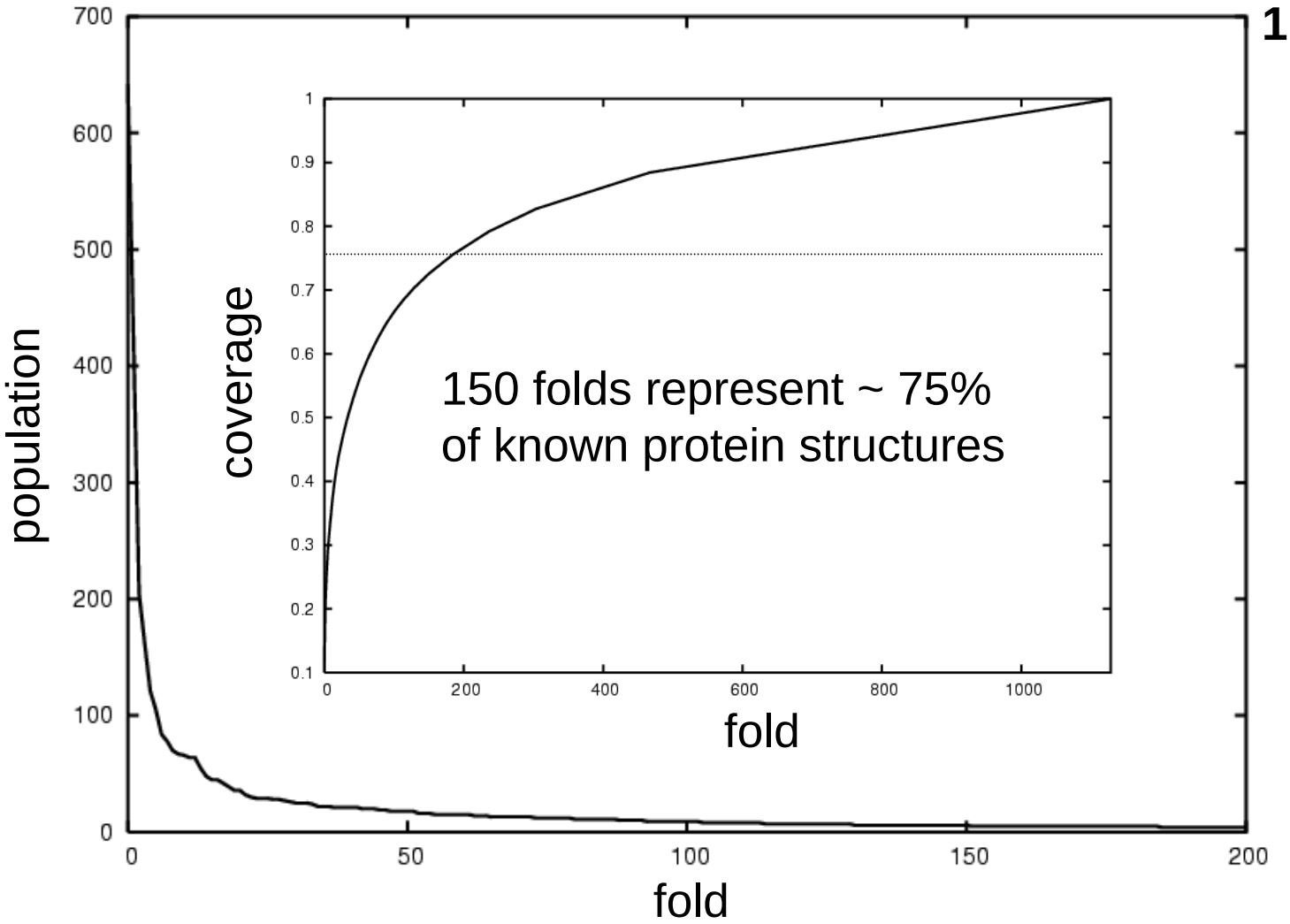
1du6



1cqt

Dynameomics

- **Simulate representative protein from all folds**



1

1. Day R., Beck D. A. C., Armen R., Daggett V. Protein Science (2003) 10: 2150-2160.

Dynameomics

- **Simulate representative protein from all folds**
  - **Native (folded) dynamics**
    - **20 nanosecond simulation at 298 Kelvin**
  - **Folding / unfolding pathway**
    - **3 x 2 ns simulations at 498 K**
    - **2 x 20 ns simulations at 498 K**
  - **Each target requires 6 simulations**

=

MANY CPU HOURS

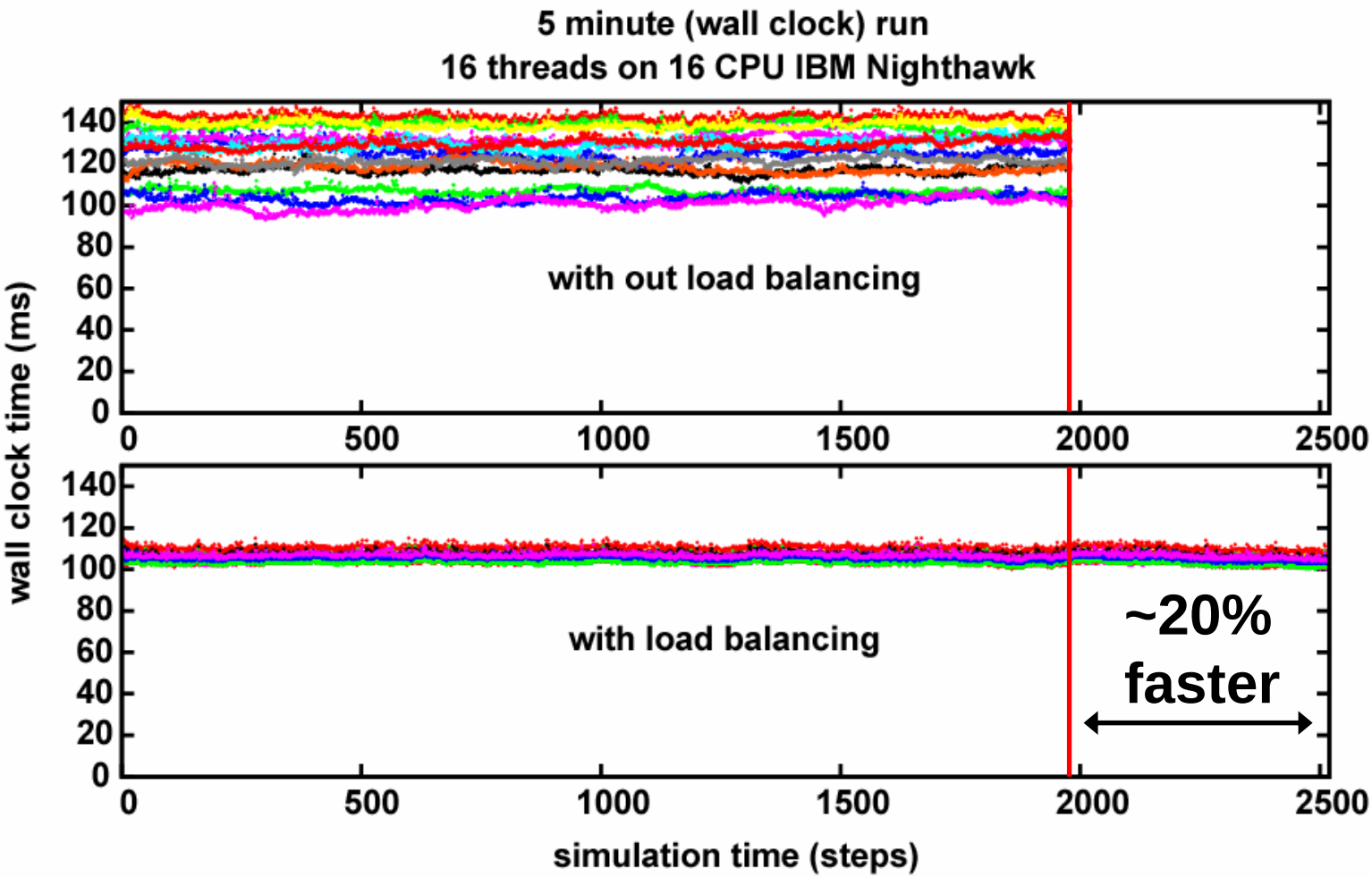
Dynameomics

- **NERSC DOE INCITE award**
  - **2,000,000 + hours**
  - **906 simulations of 151 protein folds on Seaborg**

  - **One to two simulations per node (8 – 16 CPUs / simulation)**
  - **Opportunity to tune *il*mm for maximum performance**

Dynameomics

- **Load balancing**
  - **Even distribution of non-bonded pairs to processors**



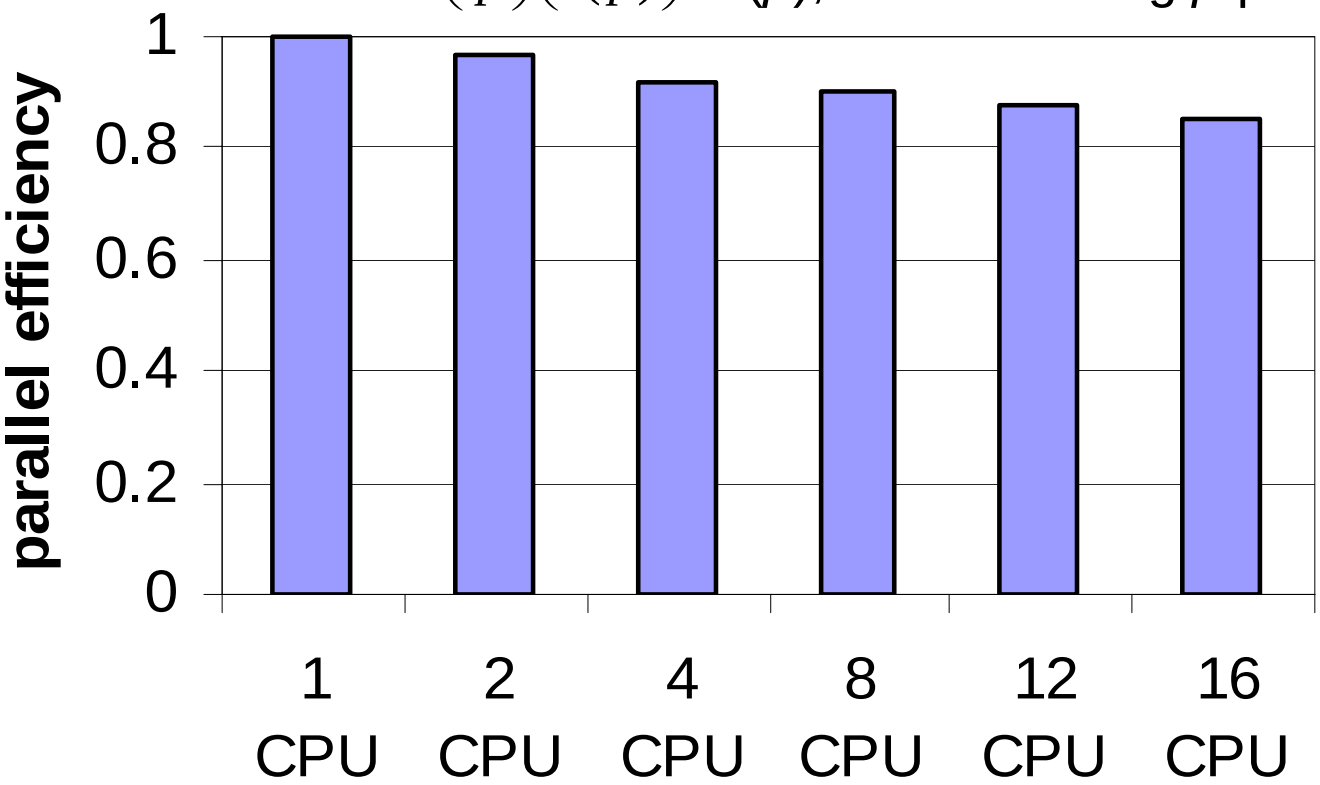
Dynameomics

- **Parallel efficiency**
  - **Threaded computations on 16 CPU IBM Nighthawk**

parallel efficiency, $e(p) = \left(\frac{1}{p}\right)\left(\frac{t(1)}{t(p)}\right)$

p, number of processors

t(*p*), run-time using *p* processors



Dynameomics

- **Simulate representative from top 151 folds**
  - 151 folds represent about 75% of known proteins
    - ~ 11 μs of combined sim. time from 906 sims!
    - ~ 2 terabytes of data (w/ 40 to 60% compression!)
    - ~ 75 / 151 have been analyzed
    - Validated against experiment where possible

Dynameomics

- **Now what?**
  - **Simulate the top 1130 folds (>90%)**
    - **More CPU time**
  - **Share simulation data from top 151 folds w/ world:**

www.dynameomics.org

    - **Coordinates, analyses, available via WWW**
    - **MicrosoftSQL database w/ On-Line Analytical Processing (OLAP)**
    - **End-user queries of coordinate data, analyses, etc.**
  - **Data mining**
    - **More CPU time, clever statistical algorithms, etc.**

Acknowledgements

- DOE / NERSC's INCITE (David Skinner, *et al*)
- NIH
- Microsoft, Inc.
- Structures rendered using Chimera, Molscript, Raster3D & PyMOL